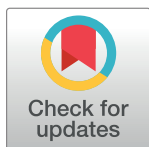


RESEARCH ARTICLE

Factor structure of intelligence and divergent thinking subtests: A registered report

Russell T. Warne ^{*}, Sam Golightly, Makai Black

Department of Behavioral Science, Utah Valley University, Orem, Utah, United States of America

^{*} russwarne@gmail.com

This is a Registered Report and may have an associated publication; please check the article page on the journal site for any related articles.

 OPEN ACCESS

Citation: Warne RT, Golightly S, Black M (2022) Factor structure of intelligence and divergent thinking subtests: A registered report. PLoS ONE 17(9): e0274921. <https://doi.org/10.1371/journal.pone.0274921>

Editor: Sergio Agnoli, University of Trieste, ITALY

Received: February 19, 2022

Accepted: September 6, 2022

Published: September 19, 2022

Copyright: © 2022 Warne et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: A public version of the data (with identifying information removed) is available at <https://osf.io/8rpfz/> DOI: [10.17605/OSF.IO/8RPFZ](https://doi.org/10.17605/OSF.IO/8RPFZ).

Funding: This study was funded by a Utah Valley University Undergraduate Research Scholarly and Creative Activities (URSCA), a Utah Valley University College of Humanities and Social Sciences Engaged Learning in the Liberal Arts (ELLA) grant, and a quick grant from the Utah

Abstract

Psychologists have investigated creativity for 70 years, and it is now seen as being an important construct, both scientifically and because of its practical value to society. However, several fundamental unresolved problems persist, including a suitable definition of creativity and the ability of psychometric tests to measure divergent thinking—an important component of creativity—in a way that aligns with theory. It is this latter point that this registered report is designed to address. We administered two divergent thinking tests (the verbal and figural versions of the Torrance Tests of Creative Thinking; TTCT) with an intelligence test (the International Cognitive Ability Resource test; ICAR). We then subjected the subscores from these tests to confirmatory factor analysis to examine which of nine theoretically plausible models best fits the data. Results show that none of the pre-registered models fit the data well, an ambiguous result that leaves unanswered the question of whether intelligence and divergent thinking tests measure the same construct. Exploratory (i.e., not pre-registered) measurement models of each test separately shows that the TTCT-F may not measure a coherent, unitary construct—leading to model misspecification when TTCT-F subtests were included in larger models. This study was conducted in accordance with all open science practices, including pre-registration, open data and syntax, and open materials (with the exception of copyrighted and confidential test stimuli). Materials are available at <https://osf.io/8rpfz/>.

Introduction

In 1950, J. P. Guilford gave his presidential address to the American Psychological Association, calling for psychologists to produce more research on creativity. Ever since Guilford's [1] address, creativity has been one of the most valued topics for educational and differential psychologists to study. Creativity in the sciences and the arts is an engine for economic and cultural progress [2] and important in its own right as a construct.

One of Guilford's [1] concerns when he encouraged more research on creativity was that psychology's emphasis on analytical intelligence in testing and research made scholars neglect creativity. In Guilford's view, "creativity and creative productivity extend well beyond the domain of intelligence," [1, p. 445], and intelligence tests were unable to measure creative methods of problem solving. Guilford theorized over the ensuing decades until his death about

Valley University Grants for Engaged Learning (GEL). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

the relationship between intelligence and creativity. By the end of his life, he had incorporated creative thinking functions into his sprawling Structure of Intellect Model [3] as one of the operations humans perform while solving cognitive tasks [4].

In discussing the relationship between creativity and intelligence, Guilford [1] stated that the two constructs were undoubtedly positively correlated. Studies designed to analyze the relationship between creativity and intelligence have consistently produced modest correlations, often in the range of $r = .10$ to $.30$ [5–9]. One of the seminal studies on creativity and intelligence was Wallach and Kogan's [10] analysis of 151 5th graders. The children were given multiple creativity and intelligence tests. Scores from the tests within each group intercorrelated well with one another ($r = .41$ between creativity tests, and $r = .51$ between intelligence tests), but the correlation between these categories was low ($r = .09$; [10]), providing strong evidence of divergent validity. Silvia [5] later used latent variable analysis and replicated Wallach and Kogan's [10] results, giving strong evidence for a consistent yet marginal relationship between creativity and intelligence. In a similar publication, Plucker [11] reanalyzed creative achievement data from Torrance's [12] longitudinal study. Using structural equation modeling, he found that divergent thinking scores predict creative achievement with much stronger predictive validity than intelligence scores and concluded that divergent thinking and intelligence "represent relatively independent constructs" [11, p. 111].

This finding is sometimes disputed, though. Methodological and assessment artifacts make the correlation between intelligence and creativity scores appear artificially weaker [13]. Silvia [9] found that latent variable correlations tend to be stronger than observed variable correlations, and that sample size, task type, and the number of responses can influence the apparent correlation between intelligence and creativity test scores. Studies based on the Berlin Structure of Intelligence model seem to indicate that some cognitive abilities that contribute to intelligence have moderate or strong relationships with divergent thinking and/or creativity, including mental speed [14], working memory [15, 16], and associative fluency [15]. This may indicate the correlation between intelligence and creativity test scores is due to an overlap in at least some cognitive abilities that contribute to behaviors that are labeled "intelligent" or "creative." And in the domain of mathematical creativity, the correlation with intelligence test scores is often stronger than the correlation between general creativity measures (like divergent thinking scores) and intelligence [7].

The mere presence of a positive correlation between creativity and intelligence test scores is not enough to establish the nature of the relationship between the two constructs. A variety of theoretical and causal models could produce the positive correlations so frequently found between measures of intelligence and creativity [6]. Before discussing these models, it is important to explore the definitions of both creativity and intelligence.

Definitions and components of creativity

Theories of creativity reach back thousands of years. Acknowledging differences in thought processes was present in the culture of Ancient Greece [17], though people also "believed that creativity required the intervention of the muses" [18, p. 152] and that artists were not considered creative in and of themselves. The earliest empirical inquiry into human creative behavior can be traced to Galton's [19] work on human capabilities. In more recent decades, defining creativity has been perhaps the most central topic and challenge in contemporary creativity research [20]. Guilford [1] was an early champion of the importance of scientific research in creativity, yet even he seemed to struggle defining what creativity is; "In its narrow sense, creativity refers to the abilities that are most characteristic of creative people" (p. 444). Runco and Jaeger [20] contended that the first clear scholarly description of creativity was offered by Stein

[21], who defined creativity as, “a novel work that is accepted as tenable or useful or satisfying by a group in some point in time” (p. 311). Stein’s [21] definition has been adopted by many researchers [22–24] and includes two important criteria: originality and usefulness.

Originality refers to an idea or product not having previously existed. This is crucial to the construct of creativity. The ability to convincingly forge the Mona Lisa may require talent, but it does not necessarily mean a person is creative. Creativity requires novelty. However, many “ideas and products that are merely original might very well be useless” [20, p. 93], and so novelty is a necessary but not sufficient qualification for creativity. Therefore, for products to qualify as creative, they must be original *and* useful.

The dual criteria of originality and usefulness have been troublesome for creativity researchers, primarily because judging originality and usefulness involves some subjectivity [25, 26]. The novelty required for a product to qualify as creative could theoretically be measured objectively, but this would require accurately determining whether a product has previously existed somewhere in the world at some point in time. To prove that an idea is truly novel and has never existed before is inherently illogical because it requires proving a negative. Usefulness is even more subjective because it relies almost entirely on context. This usefulness is often determined by the community or group that the product is created for or by individuals who encounter the idea or product. Runco et al. [27] postulated that from this perspective, “creativity can lead to things which are original and useful but only for the individual creator himself or herself” [27, p. 366]. On the other hand, Corazza’s [28] dynamic definition of creativity, recognizes that value judgements of usefulness can be mistaken and is useful for understanding creativity as a process that occurs within an environmental, cultural, and individual context. Thus, the subjectivity of judgements of creative accomplishments is not a problem in this viewpoint and is simply part of the challenge of bringing creative ideas to fruition—and a topic worthy of study in its own right.

Non-cognitive influences on creativity

Although Stein’s [21] requirements that creative thought results in original and useful products, others have expanded theories of creativity to include other behaviors and characteristics. Another important facet of creativity is the influence of non-cognitive elements like personality and motivation on its expression [22]. For example, Furnham and Bachtiar [29] and Furnham and Nederstrom [30] found that personality trait extroversion correlated positively with test scores of divergent thinking. Similarly, Wang et al. [31] found that extroversion correlated substantially with scholarly creative achievement in undergraduate students. Openness to experience has also been shown to relate to creative expression. Benedek, et al. [32] found that jazz musicians scored higher on tests of divergent thinking and openness to experience compared to their folk and classically trained counterparts. Fink and Woschnjak [33] found a similar pattern in dancers; modern/contemporary performers scored higher in creativity and openness to experience than individuals trained in theater and ballet. In addition to the direct relationship between openness and creativity, openness mediates the relationships between temperament variables and creativity [34]. Temperament can also directly influence creative expression. For example, Nęcka and Hlawacz [35] found the temperament trait activity correlated positively with creativity, while emotional reactivity was negatively correlated.

Motivation also likely influences creative expression. Hannam and Narayan [36] found that university students were more likely to produce creative work if they were intrinsically motivated and perceived the work environment as fair. Saether [37] also discovered that motivation mediated between creativity and fairness in a sample of online survey responses. These results seem in line with Silvia et al.’s [38] assertion that motivation and its effect on creativity

includes aspects of, “goals, self-regulatory processes, and experiences that foster or impede wanting to invest time in creative activities” (p. 114).

Moreover, it is possible that many individuals possess the capacity to be creative but choose not to do so [23]. Torrance acknowledged that scores from creativity tests do not ensure creative accomplishment [27]. Because of its multifaceted nature, with roots in personality, the social environment, and cognition, Mumford and Gustavson [39] have argued that creativity ought to be considered the product of a system of characteristics, rather than a single, isolatable characteristic.

Thus, creativity is a complex, multifaceted construct. However, there are commonalities among the dueling definitions. Weiss et al. [13] found that, across definitions, idea generation (often called fluency) and originality are part of nearly every scholarly definition of creativity. These two aspects of creative behavior are well captured in the concept of divergent thinking, which is the focus of our scholarly investigation.

Importance of creativity

Creativity represents one of the pinnacles of human experience. Indeed, it is difficult to imagine human progress in any of its forms without the catalyzing spark and sustaining force of creativity. Vygotsky held that, “creativity is an essential condition for existence and all that goes beyond the rut of routine” [40, p. 11]. This alone makes creativity worthy of research, but some theorists have proposed other reasons.

One reason to empirically study creativity is its role in fostering economic growth. There is a major economic need for domestic and international industries to quickly and efficiently solve problems, and many simple jobs that need little to no creative output are being automated [41]. It will become increasingly important for the advancing global economy to foster and identify creativity in all age groups [39]. Creativity is crucial in educational organizations as well. In a review of the empirical creativity literature, Davies et al. [42] found evidence that creative learning environments positively impacted student academic attainment, concentration, enjoyment, enthusiasm, and emotional development.

Measuring creativity with the torrance tests of creative thinking

An essential requirement for the empirical study of creativity is a method of measuring creative behavior. The most widely used measures of creativity are the Torrance Tests of Creative Thinking (TTCT), which consist of a Verbal test reliant on written linguistic responses and a non-verbal Figural test that uses pictorial stimuli and requires responses that the examinee must draw (see detailed description below). However, the TTCT does not measure the totality of creative thinking; instead, it measures one of the building blocks of creative behavior: divergent thinking, i.e., the capacity to produce a variety of ideas in response to a stimulus [43]. The TTCT measures divergent thinking through fluency (i.e., the number of responses generated), originality (i.e., how much the responses differ from common responses given in the norm sample), and flexibility (which is the variety of types of responses that examinees give). Divergent thinking has been shown to positively correlate with individual creative achievement, most notably in a 50-year longitudinal study where TTCT fluency, flexibility, originality, and elaboration scores all correlated $r = .20$ to $.29$ with personal achievements 50 years later, whereas IQ scores had a slight negative correlation with later achievement. [27] Indeed, divergent thinking is so important to creativity that some researchers claim that divergent thinking is the most valid way to predict creativity, almost using the two terms interchangeably [44]. However, others deny these claims and hold that divergent thinking is only a predictor of individual creativity [29, 45].

Despite the popularity of the TTCT, there remain uncertainties about some of the tests' psychometric properties. One of the questions regarding the TTCT is the dimensionality and factor structure of its scores. Multiple studies have suggested that the TTCT-F and TTCT-V are multidimensional and follow a two-factor model [46–50]. The authors of each of these studies termed these two factors innovative and adaptive, patterned after Kirton's [51] Adaptor-Innovator theory (KAI). The KAI postulates problem-solving and creativity are often manifested in one of two styles; the adaptive style is characterized by doing things better, while the innovator style is characterized by doing things differently [51]. In each of these studies fluency and originality loaded onto the innovative factor, and elaboration and abstractness of titles loaded onto the adaptive factor [46–50]. Resistance to premature closure had less consistent results, loading onto both the innovative and adaptive factors in [46] and [49], just the innovative factor in [48], and just the adaptive factor in Humble et al. [47] and [50].

If the KAI theory is correct, it would present a problem for TTCT because the tests' subscores and global score do not align with the theory. Instead of two subscores—innovation and adaptation—each version of the TTCT produces a global score and three (for the TTCT-V) or six (for the TTCT-F) subscores. This may represent a major deficit in the internal validity evidence of the TTCT and its ability to give test users a correct understanding of examinees' divergent thinking. Because our study is psychometric in nature, it will not provide definite answers to the question of how creativity and intelligence relate to one another as constructs. Instead, this study is limited to investigating the degree to which divergent thinking and intelligence subtests form separate latent factors.

In summary, one of the intransigent problems with the TTCT is that its factor structure (1) may not align with leading theories of the nature of creativity, and (2) the interpretation of its scores—especially the global scores that the TTCT produces. This means that the construct validity and internal validity of the TTCT-F and TTCT-V are disputed. An investigation into the factor structure of the tests could clarify the psychometric properties of both versions of the TTCT.

Defining intelligence

Like most constructs in the social sciences, there are many definitions of intelligence (see [52] for a compilation of definitions from leading theorists). Since the mid-1990s, though, one conceptual definition has found widespread—though not unanimous—agreement among experts in intelligence:

Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—“catching on,” “making sense” of things, or “figuring out” what to do. [53, p. 13]

Most definitions overlap with this one, often encompassing a global ability to engage in problem solving (e.g., [54]) or learn from one's environment and/or experience (e.g., [55]), though some theorists and researchers have proposed definitions that are neurological in origin (e.g., [56]) or culturally specific (e.g., [57]), or that extend beyond cognitive abilities (e.g., [58]). Alternative definitions, however, have not found widespread acceptance, and most psychologists still consider complex reasoning and general cognitive competence as some of the central components of intelligence.

In addition to verbal definitions, many psychologists subscribe to a statistical definition, where intelligence is taken as being similar or equivalent to a general factor that statistically captures approximately half of variance in the scores on a set of cognitive tasks. This factor is often called *g* or Spearman's *g* (in honor of its discoverer). Unlike a verbal definition, the statistical definition of intelligence is much less subjective or ambiguous. Moreover, because of the indifference of the indicator, the *g* factor that emerges from different intelligence tests are nearly identical (with factor correlations often $r = .95$ or higher), indicating that *g* as a statistical definition of intelligence is not dependent on any particular test or collection of tasks [59–65]. In this paper, we will subscribe to the statistical definition and use a common factor of scores on cognitive tasks as our operationalization of intelligence.

Investigating the relationship between divergent thinking subtests and intelligence subtests via factor structure

Factor structure of intelligence tests

In contrast with the TTCT, the factor structure of scores from intelligence tests is well established. Almost all cognitive test batteries produce a series of scores that either load onto a single factor or produce a number of factors that, in turn, load onto a single general factor [65, 66]. This general factor has been called *g* since Spearman [67] discovered it over a century ago. Among psychologists studying intelligence, *g* is a mainstream theory, and there is strong evidence that every cognitive task loads on *g* to some extent [64]. Spearman [68, pp. 197–198] called this phenomenon the “indifference of the indicator,” and it has led some experts to argue that every task in life that requires cognitive work is its own intelligence test, which would explain why IQ scores correlate with so many life outcomes [69–71].

The indifference of the indicator presents a major challenge to the widely accepted belief that creativity tests—especially the TTCT with its focus on divergent thinking—measure a distinct construct from intelligence. TTCT test content is unquestionably cognitive, and according to the theory of the indifference of the indicator, TTCT content should measure *g*, at least partially. Moreover, in the past, some researchers designing tests to measure other constructs have accidentally created tests that measured intelligence [64, Chapter 7]. For example, Sanders et al. [72] found that the Defining Issues Test, a test designed to measure moral development and reasoning, is *g*-loaded and is a moderately good measure of verbal intelligence. Likewise, literacy tests are highly *g* loaded and—in American samples—do not seem to measure a construct that is distinct from intelligence [73]. These examples and the strong evidence in favor of the indifference of the indicator raise the possibility that the TTCT is actually a measure of intelligence. The best way to investigate this possibility is through psychometric study of the factor structure of scores from the TTCT and an intelligence test when given to the same sample of examinees.

The factor structures of both the TTCT and intelligence tests have been studied independently, but we have been unable to identify any research examining the factor structure of creativity and intelligence tests at the same time to determine whether these tests measure the same latent construct, or multiple constructs (and how multiple constructs might be related to one another). Our goal in conducting this study is to determine the factor structure of subtests drawn from intelligence and divergent thinking tests when administered together. Through this study we aimed to determine the degree to which divergent thinking subtests and intelligence subtests measures separate constructs. We used confirmatory factor analysis to determine which of a series of plausible factor structures provides the best fit for the subtest data.

Theoretically plausible factor structures

Sternberg and O'Hara [74] described five possible ways that could describe the relationship between intelligence and creativity:

- Creativity is a component of intelligence.
- Intelligence is a component of creativity.
- Creativity and intelligence are different constructs with overlapping components and/or mental processes.
- Creativity and intelligence are different labels for what are, substantially, the same problem-solving construct.
- Creativity and intelligence are separate constructs, with any correlations between the two being incidental.

Sternberg and O'Hara [74] found that there was evidence supporting all of these viewpoints, making it hard to distinguish which one is the best description of the relationship between intelligence and creativity. Years later, Karwowski et al. [6] added another possibility: that intelligence is a necessary but not sufficient condition for high creativity, which may explain the positive correlations between the two but also why many high-IQ individuals fail to engage in creative behaviors. Complicating the picture is that the relationship between the two constructs may be dependent on the context, [75] with a degree of domain knowledge often being necessary for a person to generate creative ideas [2].

With the exact nature of the relationship between intelligence and creativity remaining an unresolved question, we posited that the positive correlation between the two constructs may be partially due to a measurement artifact that some tasks that measure creativity may be—like all other cognitive tasks—partially measuring intelligence. To investigate this possibility, we conducted this study to determine whether the divergent thinking tasks on the TTCT are measures of general intelligence (i.e., are *g*-loaded). Such a study would help psychologists in both fields (i.e., creativity and intelligence) determine whether the apparent relationship between scores on tests measuring intelligence and creativity is due to a measurement artifact or to a real relationship between the two constructs. Understanding the nature of the correlation between intelligence and creativity would help clarify (and possibly support or falsify) theories of the relationships between the two constructs and improve research in future studies.

The possibility that divergent thinking tests also measure intelligence is not far-fetched, and the history of intelligence testing has several examples of measures of divergent thinking being interpreted as measuring intelligence. Over 100 years ago, Binet and Simon [76, pp. 229–230] included a task on their second intelligence test in which a child examinee was asked to name as many words as possible in three minutes. Although the task was scored quantitatively by tallying up the total number of words that the examinee produced, Binet discussed the qualitative differences in responses among children, noting the different categories of words that some examinees generated or the uniqueness of their responses. Modern creativity researchers would recognize Binet's quantitative scoring procedure as a measure of fluency and his qualitative analysis as touching upon originality and flexibility in responses [43]. More recently, Carroll [66] identified fluency as a manifestation of a broad mental retrieval ability that was subsumed by a general intelligence factor. Other researchers from the psychometric tradition have also suggested that common measures of creativity and/or divergent thinking could be measuring aspects of intelligence [13].

From a psychometric perspective, there are several factor structures that we believe are plausible for explaining the relationship between scores on divergent thinking and intelligence

tests: (a) distinct but correlated constructs, (b) scores loading directly on a *g* factor, and (c) a hierarchical structure with *g* as a second-order factor. We will briefly describe each of these possibilities here and then describe exact models we will use in our study in the Methods section.

Possibility (a), which is a factor structure of correlated separate constructs, would emerge if intelligence and divergent thinking subtests produce separate factors. This is the factor structure suggested by divergent validity research showing that intelligence and divergent thinking tests do not strongly correlate with one another (e.g., [11]). In the most straightforward form, this could occur if the TTCT and intelligence tests each produced their own factor(s) which were, in turn, correlated with one another. Structure (b) would occur if subscores from the divergent thinking subtests and intelligence tests all formed a single, undifferentiated general factor. This would support the intelligence community's belief that all cognitive tasks measure *g* to some extent and be strong evidence that Spearman's [68] theory of the indifference of the indicator is correct. Finally, possibility (c) would be a hybrid model of (a) and (b) where a number of factors can combine to form a single second-order *g* factor. This type of structure would support intelligence theorists' belief in the universality of *g* but also permit the existence of first-order factors—such as divergent thinking and intelligence factors that span a number of subscores. In this study, we assessed 1–4 measurement models of each type in order to better understand how intelligence test scores and divergent thinking test scores are related.

Methods

Ethics

The study received ethical approval from the Utah Valley University Institutional Review Board, protocol #441. Informed consent was obtained at two time points: when examinees took the cognitive ability test online and again at the beginning of the in-person testing session (when the TTCT tests were administered). The initial informed consent was recorded online, while the informed consent obtained at the beginning of the in-person testing session was obtained through a signed written consent document.

Instruments

We collected a total of thirteen subscores drawn from three professionally developed psychometric tests: the Torrance Tests of Creative Thinking Figural Test A, the Torrance Tests of Creative Thinking Verbal Test A, and an abbreviated version of the International Cognitive Ability Resource (ICAR) test.

TTCT figural test A. The TTCT Figural Test A consists of a picture construction activity, picture completion activity, and lines activity. The test is designed to measure divergent thinking with standardized pictorial stimuli.

The scoring system produces six subscores:

- fluency (defined as the capacity to produce a large number of visual images),
- originality (the production of unusual responses),
- elaboration (the capacity to create responses that are more embellished than a basic figure),
- abstractness of titles (the capability of producing non-literal titles for pictures),
- resistance to premature closure (generating responses that leave stimuli open-ended and do not close them immediately and prematurely), and
- a checklist of creative strengths.

The first five subscores are norm-referenced, and examinees receive points based on the degree to which their responses are more creative than those generated by the test's norm sample. The checklist of creative strengths subscore is created by summing 13 criterion-referenced scores that correspond to components of the examinees' constructed responses on the 3 subtests [77].

The TTCT Figural Test A subtests were administered with the time limits indicated in the test manual (10 minutes for each TTCT-F task, for a total of 30 minutes). However, we modified the instructions slightly because they seemed to be written for children, and our examinees were adults. These modifications were minor in nature and were designed to remove language that we found overly simplistic or condescending. Altered instructions are available from the project's page on the web site at <https://osf.io/8rpfz/>.

TTCT verbal test A. Like the TTCT Figural Test A, the TTCT Verbal Test A is designed to measure divergent thinking in a standardized fashion. The TTCT Verbal Test A consists of three ask-and-guess tasks, a product improvement task, an unusual uses task (very similar to the Alternate Uses Task), and a just suppose task. These six tasks generate three subscores: fluency, flexibility, and originality. Fluency and flexibility are the same as for the TTCT Figural Test A. For the TTCT Verbal Test A, the flexibility measures examinee's adeptness at generating responses in different categories.

Just as with the TTCT Figural A test, the tasks were administered with the time limits specified in the test manual (5 or 10 minutes for each TTCT-V task, for a total of 40 minutes). The only modification we made to this test was to alter the instructions to better suit an adult examinee population. Altered instructions are available from the project's page on the web site at <https://osf.io/8rpfz/>.

ICAR test. The ICAR test was developed as a free test of general cognitive ability that is available in the public domain for researchers to use [78]. The ICAR test consists of four subtests: verbal reasoning, 3D rotation, letter and number series, and matrix reasoning. The verbal reasoning subtest consists of items that use written English to ask questions about general knowledge, basic reasoning, and the relationships among words (e.g., synonyms, antonyms). The 3D rotation subtest shows a two-dimensional representation of a cube showing three sides. The examinee must then identify which of six options depicts a rotated version of the target cube. Examinees also have the option of indicating that none of the options is correct and another option to state that they do not know the answer. The letter and number subtest consists of a sequence of five letters or numbers (never both in the same test item), which the examinee must complete. Finally, the matrix reasoning subtest shows a 3 x 3 grid of geometric figures with one section replaced with a question mark. The examinee should select which of six options would complete the pattern shown in the matrix. The items on the ICAR test are typical for written intelligence tests. All item types have appeared on intelligence tests for at least 50 years and are well established methods for measuring intelligence in examinees [79].

To reduce the burden on our volunteer examinees in the allotted time, we were forced to make some adjustments to the ICAR test. The first was that we shortened the ICAR test from 60 items to 52 items: 16 items on the verbal reasoning subtest, 16 items on the 3D rotation subtest, 9 items on the letter and number series subtests, and 11 items on the matrix reasoning subtest. We shortened the test to fit all three tests into a 2.5-hour testing window. However, just two days after the study received IRB approval, permission for in-person data collection at our institution was halted. When resumed, in-person data collection sessions were limited to two hours. With IRB approval, we moved the ICAR testing online to keep the TTCT testing sessions within the allotted time. The shortened version of the test retained to keep the study as convenient as possible for participants. (Note that Stage 1 of the peer review for this registered report occurred after receiving final IRB approval and after deciding to move the ICAR to online administration.)

All ICAR items were scored as 0 for incorrect answers and 1 for correct answers, and each subscale's items were summed for an overall subscale score. Therefore, the minimum possible score for each of these subtests was 0 and the maximum possible score was the number of items on that subscale. Using internal consistency reliability values reported by Condon and Revelle [78, Table 3], these shortened subtests were expected to have Cronbach's α values of at least .70, based on estimates calculated from the Spearman-Brown prophecy formula.

Additionally, we added time limits to each shortened subtest and allotted examinees 11 minutes to complete the verbal reasoning subtest, 16 minutes to complete the 3D rotation subtest, 8 minutes to complete the letter and number series, and 13 minutes to complete the matrix reasoning subtest. We determined these time limits by administering the original ICAR test untimed to a convenience sample of college students and young adults. Subtests that took longer than desired to complete were shortened, and items were dropped quasi-randomly to ensure that any subtest's eliminated items varied in difficulty. The shortened ICAR test was pilot tested with more examinees with time limits based on the expected amount of time per item calculated from the untimed administrations of the full ICAR test. Further pilot testing of the abbreviated ICAR test indicated that most examinees were able to attempt almost every item on the shortened version of the test.

Data and scoring

For both TCTT tests, we used professional scorers (employed by the test publisher) to score the tests in order to improve data quality and reduce scoring errors and subjectivity that novice scorers are susceptible to. These scorers followed procedures in the two TTCT scoring manuals. [77] In short, higher scores on fluency subscales on both tests correspond to a larger number of valid responses. Higher originality subscores on both tests correspond to more responses that were unusual in the test's norm sample. For the TTCT-V, higher flexibility subscores indicate a larger number of categories (pre-defined in the test manual) that responses could be placed into, and higher creative strengths checklist subscores indicate a larger number of special types of responses (e.g., indicating emotion, humor, fantasy) and/or more frequent examples of these responses. For the TTCT-F, higher elaboration subscores are indicative of more detailed drawings, while higher scores on the abstract titles subscale correspond to a larger number of titles that are not literal and/or achieve higher levels of abstractness. Finally, the resistance to closure subscore is higher for examinees who do not complete stimulus figures to form small, enclosed polygons or shapes.

In analysis, we used TTCT subscores that had *not* been converted to standardized scores or age or grade percentiles. For the ICAR test, the raw scores were the number of test items correctly answered on each subtest. In accordance with our published pre-registration protocol, [80] data from examinees who left the testing setting early and did not attempt all three tests were not used in the analyses.

The scoring system for the TTCT Verbal and Figural tests is based on counts of valid responses (fluency), which are then used to produce scores for unusual responses (originality) and the number of categories responses fit into (flexibility). The TTCT Figural also uses counts to produce scores for abstract titles, resistance to closure, and the creative strengths checklist. Because achieving a high subscore on originality, elaboration, and any of the other scores is dependent on first generating a number of responses, these scores are confounded with the test's fluency scores. Over the years, researchers have proposed different methods for handling this confounding, [43, 81] and there is no one clear correct method for doing so. The method we have chosen and pre-registered is to divide each task's subscore by the fluency subscore for the same task to produce a rate per valid response (e.g., a rate of responses that qualify as

original). We then summed the subscores that measure the same aspect of divergent thinking, and used these summed rates in our analyses to create adjusted scores. These altered scores can be interpreted as a respondent's average rate of divergent responses (i.e., original responses, elaborate responses, etc.) per valid response. Note that the creative strengths checklist was not adjusted in this fashion because its score is derived across all tasks on the TTCT-F, and that it is not possible to control for fluency at the task level. Therefore, this score is still partially confounded by fluency.

Sample

Sample size was set by funding constraints but was large enough to meet guidelines for conducting confirmatory factor analysis studies. Our funding permitted the purchase of test materials for 460 examinees. The goal was to have sufficient data from at least 399 examinees after data loss due to incomplete tests, missing subscores, or other problems. This would allow every model to have at least 20 sample members per estimated path. This goal exceeded the recommended minimum sample size for convergence and parameter estimate stability in confirmatory factor analysis [82]. Statistical power for the chi-squared difference tests could not be calculated because realistic parameter estimates require plausible a priori estimates of all other model parameters [83], but because this is the first confirmatory factor analysis of both intelligence and creativity subscores, these estimates are not available.

Of the 519 participants who took the ICAR, 427 later took the TTCT and had analyzable data. Fig 1 shows the number of people who left the study at each stage and/or the reasons data were unanalyzable. By far, the most common reason for subject loss was that participants who completed the ICAR online did not show up at a TTCT testing session at a later date ($n = 82$). One participant participated in a TTCT data collection session but did not complete it. Of

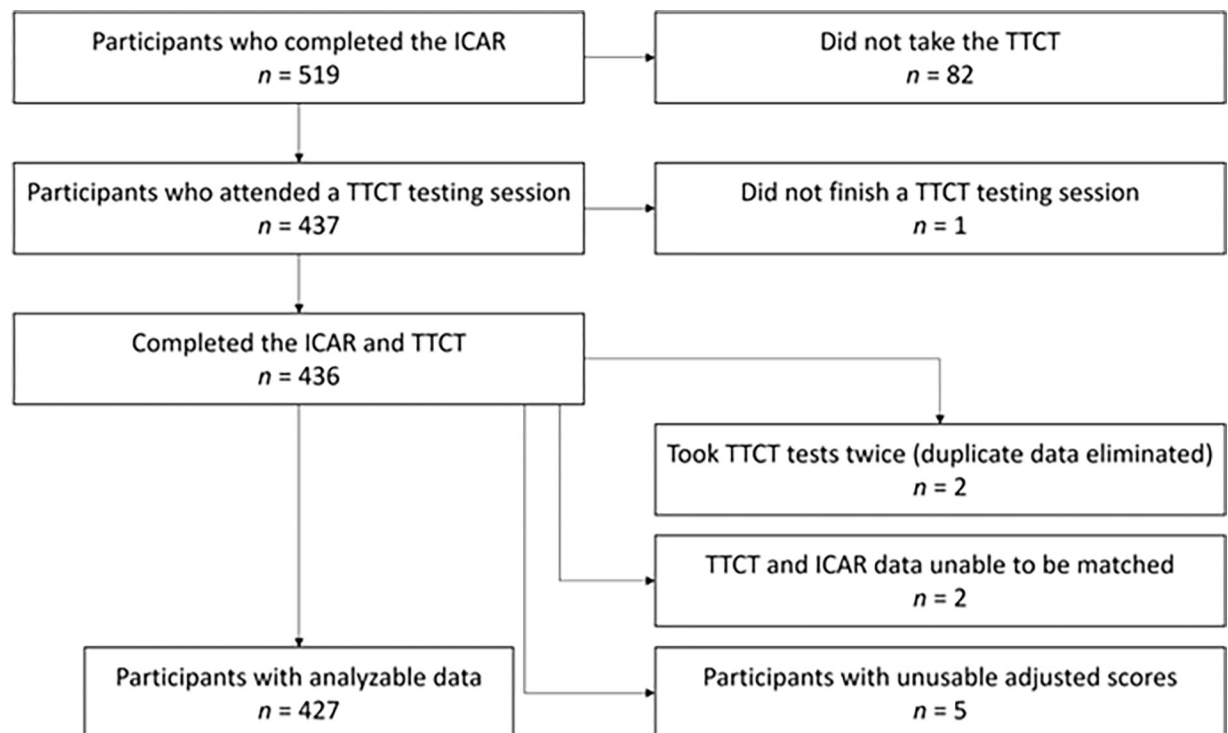


Fig 1. Subject flow diagram for the study.

<https://doi.org/10.1371/journal.pone.0274921.g001>

those who took the TTCT and ICAR, two participants took the TTCT tests twice, and data from one of their two sessions were randomly selected to be merged with their ICAR data. (The data from the other testing session were dropped from the study.) Two other participants' ICAR and TTCT data could not be merged because names and/or student ID numbers on the ICAR could not be matched to corresponding information on the TTCT tests. These two individuals were eliminated from the study. Finally, five participants had unusable adjusted scores because they had a zero fluency score on one more TTCT-F or TTCT-V task, which resulted in an undefined value for their adjusted score. Their data were also not analyzed in this study.

The sample consisted of a convenience sample of students attending a large open-enrollment university in the western United States and members of the surrounding community. Students received course credit for participating in the testing session or for inviting a community member who participated. Psychology students at this university are required to participate in research as part of their education. Examinees took the ICAR online (its typical administration format) up to three weeks before their in-person testing session. Total testing lasted 120–150 minutes (including breaks between tests), with the TTCT tests being administered in person in a session that lasted up to 110 minutes. TTCT test order was randomized for each session.

Analysis

We calculated basic descriptive statistics for all variables: means and standard deviations for all subtest scores and demographic variables that are interval- or ratio-level data, along with frequency tables for nominal- and ordinal-level demographic variables. We also created a correlation table for all subtest scores.

Our plan was to use confirmatory factor analysis to examine nine plausible models for how the subscores on the TTCT tests and the ICAR test could relate to one another. These models are summarized briefly in [Table 1](#). All confirmatory factor analyses were performed with MPlus 8.4. We believe that a Registered Report fits our study design and analysis plan because “Registered Reports can be an empowering venue for testing new theories or arbitrating between competing theories” [84, p. 570]. Having locked-in and pre-registered models removed any subjectivity from our study and ensured that the model selection process occurred without regard for our preference for any particular model.

The nine models fall into the three groups described in the introduction section and are all either hierarchical or non-hierarchical models. Models 1a, 2a, 3a, and 4a all represent a multi-factor relationship between creativity and intelligence subscores, where all factors are intercorrelated. Model 5 is a model where all scores load directly onto a general factor. The hybrid models are Models 1b, 2b, 3b, and 4b, which all have 2–4 initial factors which then load on a general factor, making them hierarchical models. All other models (i.e., 1a, 2a, 3a, 4a, and 5) are non-hierarchical models. Almost all non-hierarchical models have a corresponding hierarchical model with the same number of factors, but with the addition of a general second-order *g* factor. The exception to this is Model 5, which is a congeneric model where all subscores load directly onto a general factor.

Models 1a and 1b have three first-order factors that are based on the tests, with both forms of the TTCT and the ICAR each forming its own factor based on its subscores. Models 2a and 2b are similar, but form two first-order factors: one for all TTCT subscores and another for ICAR subscores. These models would be appropriate if both forms of the TTCT measure divergent thinking and the ICAR measures a separate cognitive ability (i.e., intelligence). Models 3a and 3b have two second-order factors based on the subtest stimuli, with verbal scores all forming a factor and non-verbal scores forming a separate, correlated factor. These models would support the traditional dichotomization between these types of stimuli on intelligence

Table 1. Proposed models for interrelationships among TTCT and ICAR subscores.

Model No.	Hierarchical?	Nested?	Reference Variable(s)	Notes
1a	No	Yes, within Models 2a and 5, and Model 4a is nested within this model	TTCT Figural Fluency, TTCT Verbal Fluency, ICAR Matrix Reasoning	Test-based model where the three tests form three correlated factors
2a	No	Yes, within Model 5, and Models 1a and 4a are nested within this model	TTCT Figural Fluency, ICAR Matrix Reasoning	Test-based model where the TTCT and ICAR form two correlated factors
3a	No	Yes, within Model 5	TTCT Figural Fluency, TTCT Verbal Fluency	Model with two correlated factors based on subtest stimuli; verbal subtests form one factor and non-verbal subtests form the second factor
4a	No	Yes, within Model 1a, which is nested within Model 2a, which is nested within Model 5	TTCT Figural Fluency, TTCT Closure Resistance, TTCT Verbal Fluency, and ICAR Matrix Reasoning	Model that splits the TTCT Figural factor in Model 2a into innovative and adaptive factors
5	No	Yes, Models 1a-4a are nested within this model (note that Model 4a is nested within Model 1a, which is nested within Model 2a, which is nested within Model 5)	ICAR Matrix Reasoning	Congeneric model with all subscores loading directly onto a single general factor
1b	Yes	No	TTCT Figural Fluency, TTCT Verbal Fluency, ICAR Matrix Reasoning, ICAR factor	Model 1a with a general factor subsuming the three test factors
2b	Yes	No	TTCT Figural Fluency, ICAR Matrix Reasoning, ICAR factor	Model 2a with a general factor subsuming the two test factors
3b	Yes	No	TTCT Figural Fluency, TTCT Verbal Fluency, Non-verbal factor	Model 3a with a general factor subsuming the verbal and non-verbal factors
4b	Yes	No	TTCT Figural Fluency, TTCT Closure Resistance, TTCT Verbal Fluency, and ICAR Matrix Reasoning	Model that splits the TTCT Figural factor in Model 2b into innovative and adaptive factors

<https://doi.org/10.1371/journal.pone.0274921.t001>

tests, which dates back to David Wechsler. Models 4a and 4b are based on empirical evidence that subscores on the TTCT Figural test do not fit a one-factor model [48, 50] and instead are best represented with a two-factor model comprising of an “innovative factor” and an “adaptive” factor, which would support Kirton’s [51] adaptor-innovator theory. (Another measure of divergent thinking showed a similar two-factor structure; see [13]). All nine models are diagrammed in Figs 2–10. There were no attempts to modify models in order to improve fit. However, because prior research shows that the creative strengths subscore on the TTCT Figural test can lead to poor model fit, [48] we also tested these models without the creative strengths subscore. These results are relegated to a supplemental file because the TTCT’s creator saw the creative strengths subscore as being essential for understanding a person’s performance on the TTCT Figural test [50].

Models were identified with the reference variable strategy, where one variable’s factor loading is set to 1.0. The fluency scores and the matrix reasoning score were used for this purpose because these scores tend to have the strongest loadings on TTCT and intelligence factors in exploratory factor analyses [85, 86]. For hierarchical models, the factor loading for the first-order ICAR factor on the second-order *g* factor was set to 1.0 because if there is a general construct that parsimoniously explains performance on the subtests, then it is likely a lower-level general intelligence factor should have a strong loading from a first-order intelligence factor.

Interpretation

Fit statistics were used to evaluate model fit and compare models with one another. We used the chi-squared value, the comparative fit index (CFI), Tucker-Lewis Index (TLI), root mean

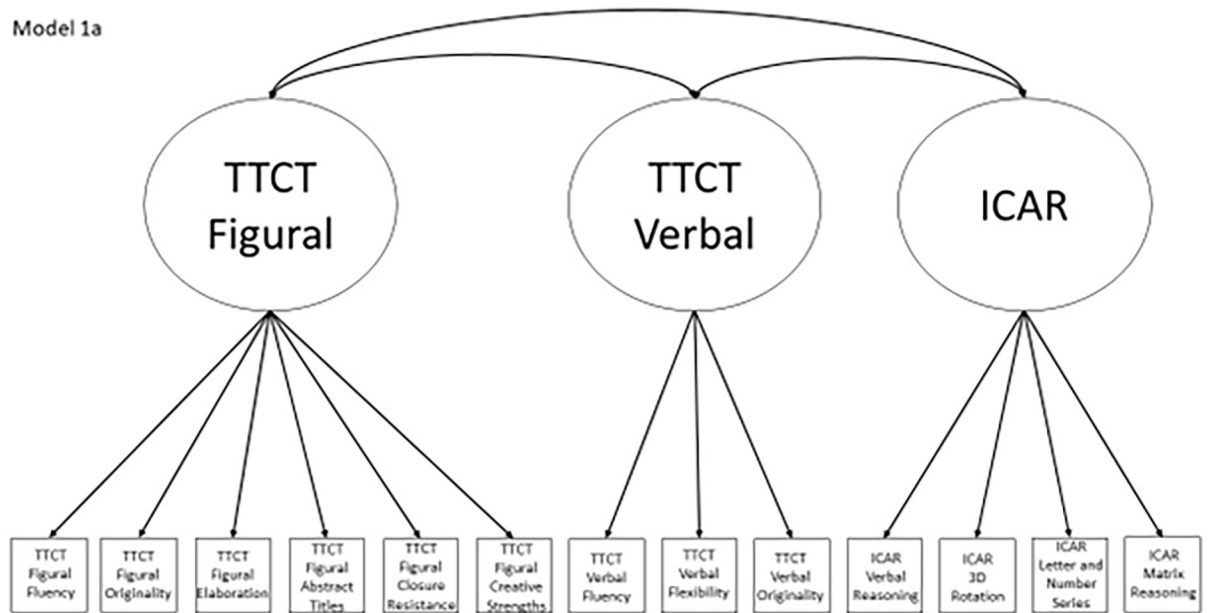


Fig 2. Confirmatory factor analysis diagram for model 1a.

<https://doi.org/10.1371/journal.pone.0274921.g002>

square error of approximation (RMSEA) with 90% confidence interval, standardized root mean square residual (SRMR), Akaike information criterion (AIC), and Bayesian information criterion (BIC). These fit indices are suitable for making comparisons among competing models, and they are a cross-section of fit statistics with compensatory strengths and weaknesses [87, 88]. For this study, models were judged to have acceptable fit if they have an SRMR value $\leq .08$ and at least one of the following: (1) CFI $\geq .90$, (2) TLI $\geq .90$, or (3) RMSEA $\leq .08$. These statistics were used to judge the best fitting model(s) among the nine. Models were

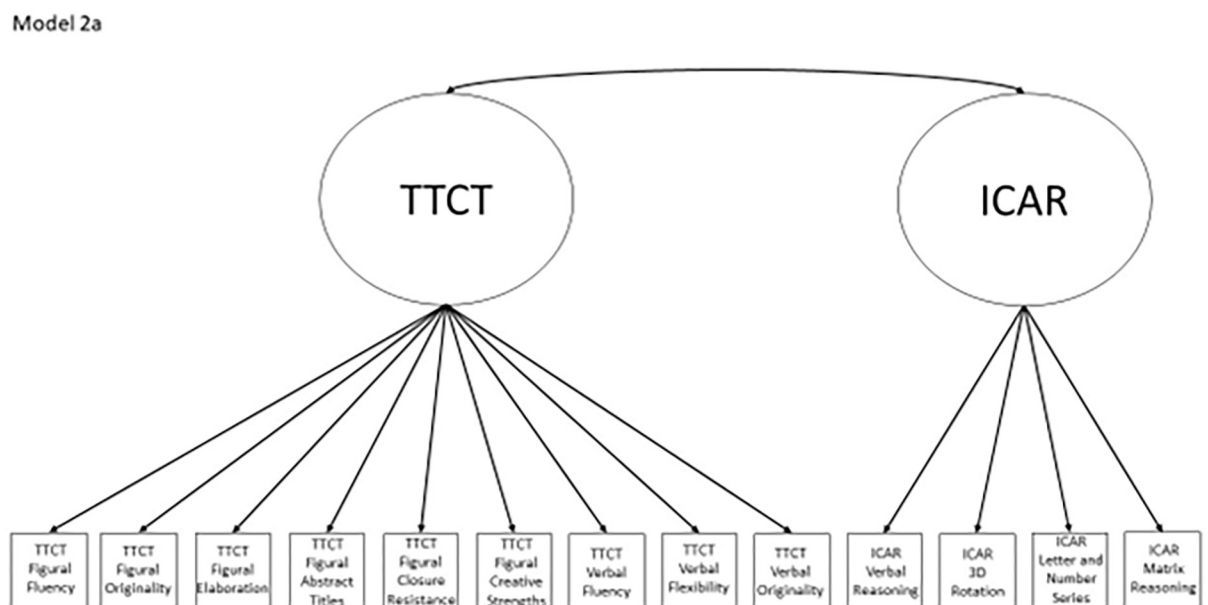


Fig 3. Confirmatory factor analysis diagram for model 2a.

<https://doi.org/10.1371/journal.pone.0274921.g003>

Model 3a

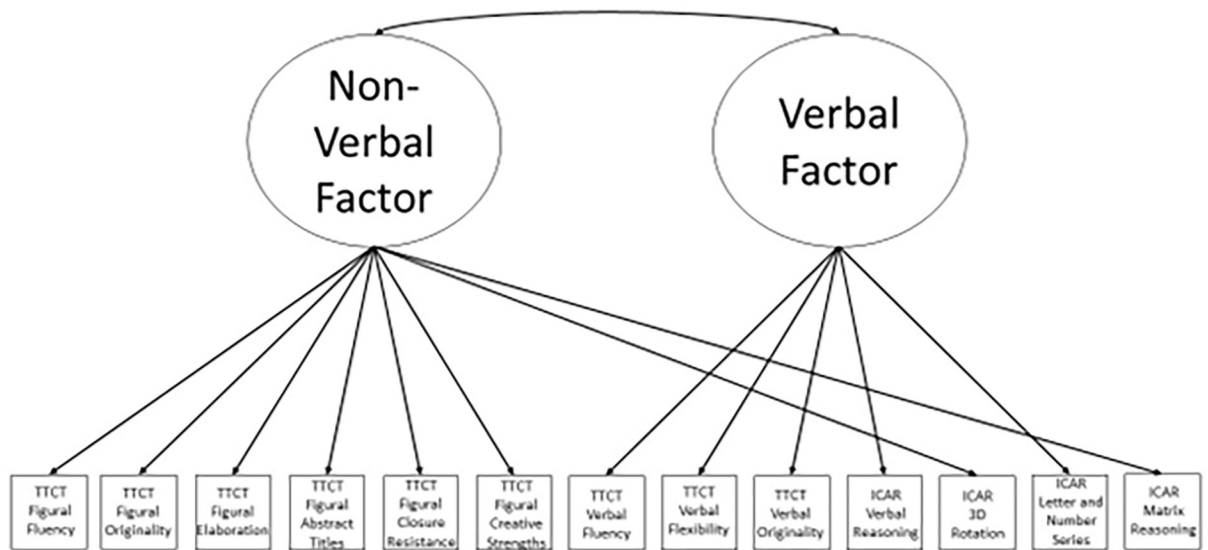


Fig 4. Confirmatory factor analysis diagram for model 3a.

<https://doi.org/10.1371/journal.pone.0274921.g004>

avored when their SRMR and RMSEA values are closer to zero and their CFI and TLI values are closer to 1.0.

The AIC and BIC were used to compare all models to one another, with lower values of these statistics indicating better model fit. When comparing models with the same degrees of freedom, both fit statistics will favor the same model, and the model with the lowest AIC and BIC will be preferred. When models have differing degrees of freedom, the penalty for a more

Model 4a

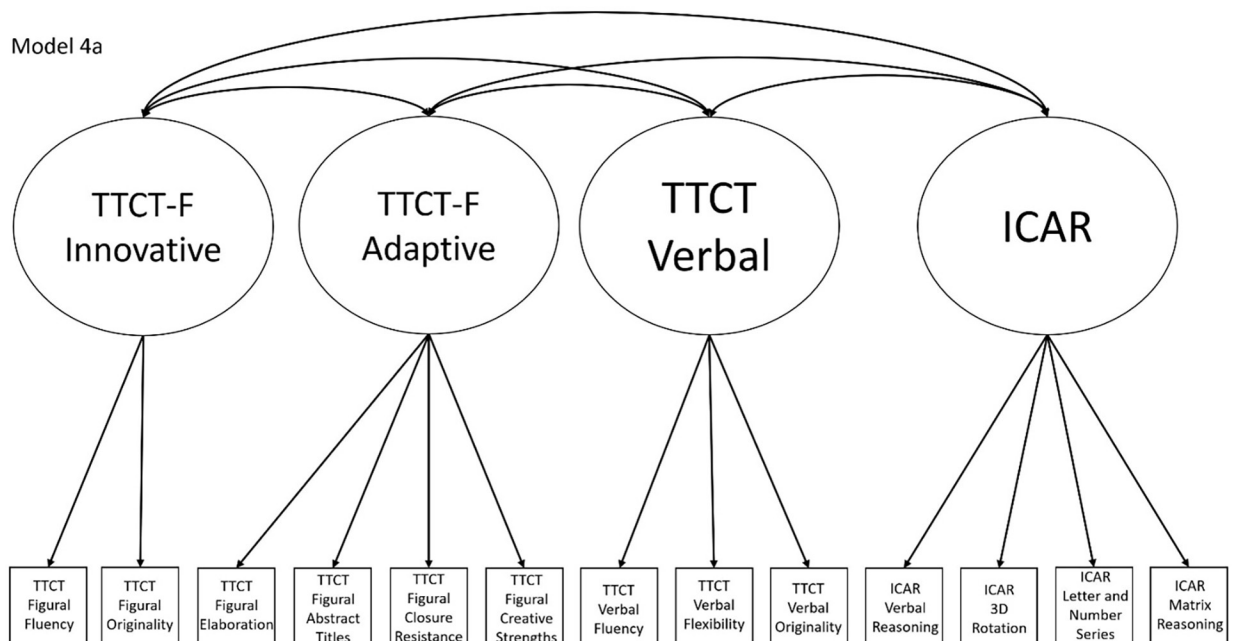


Fig 5. Confirmatory factor analysis diagram for model 4a.

<https://doi.org/10.1371/journal.pone.0274921.g005>

Model 5

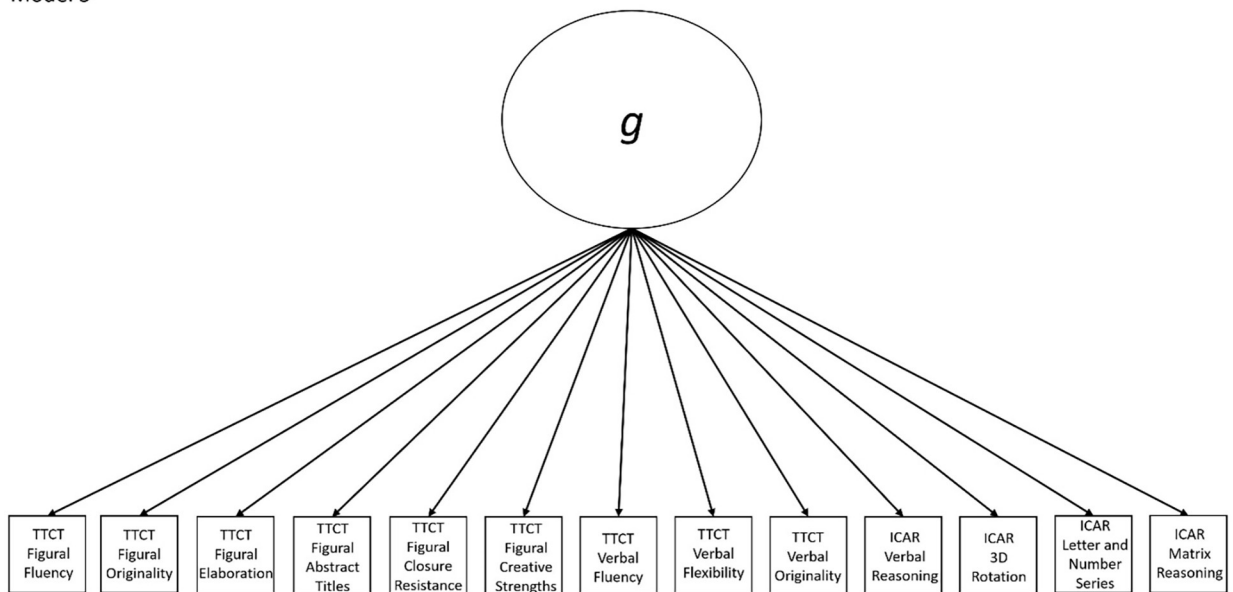


Fig 6. Confirmatory factor analysis diagram for model 5.

<https://doi.org/10.1371/journal.pone.0274921.g006>

complex model (i.e., with fewer degrees of freedom) will be more severe for the BIC than the AIC. Therefore, if a more complex model has a lower BIC than a simpler model, then it would be preferred because this would indicate that the model fits the data much better than the simpler model, despite the loss of parsimony.

Chi-squared difference tests were used to compare well-fitting nested models. Table 1 indicates which models are nested within one another. Model 5 is the most general of these

Model 1b

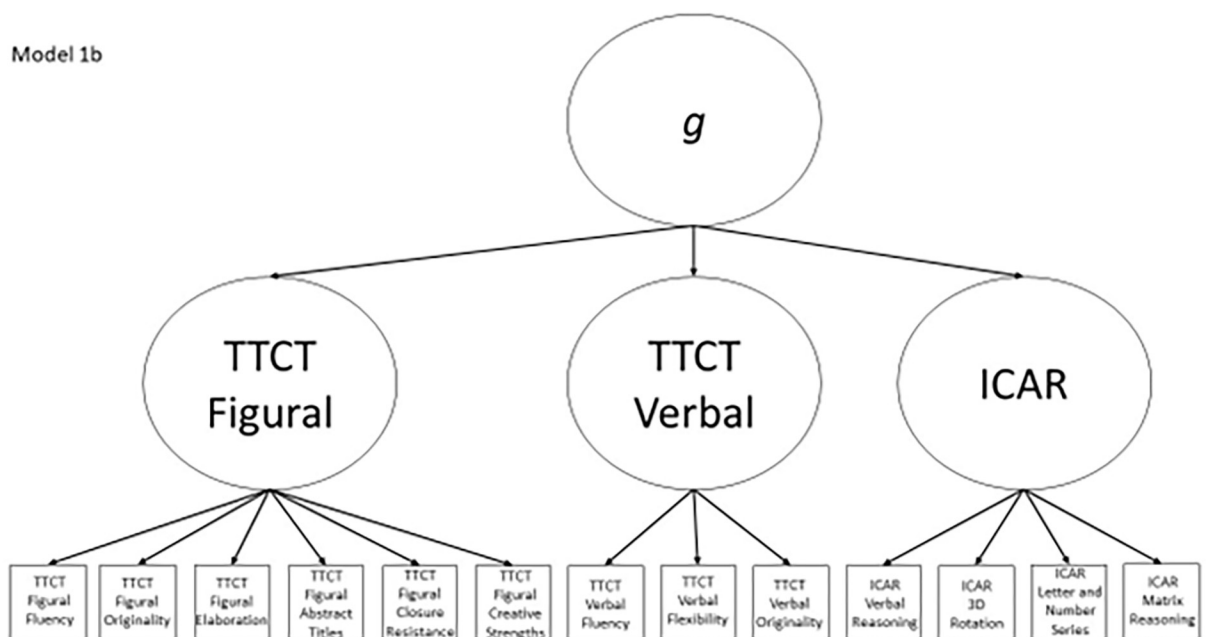


Fig 7. Confirmatory factor analysis diagram for model 1b.

<https://doi.org/10.1371/journal.pone.0274921.g007>

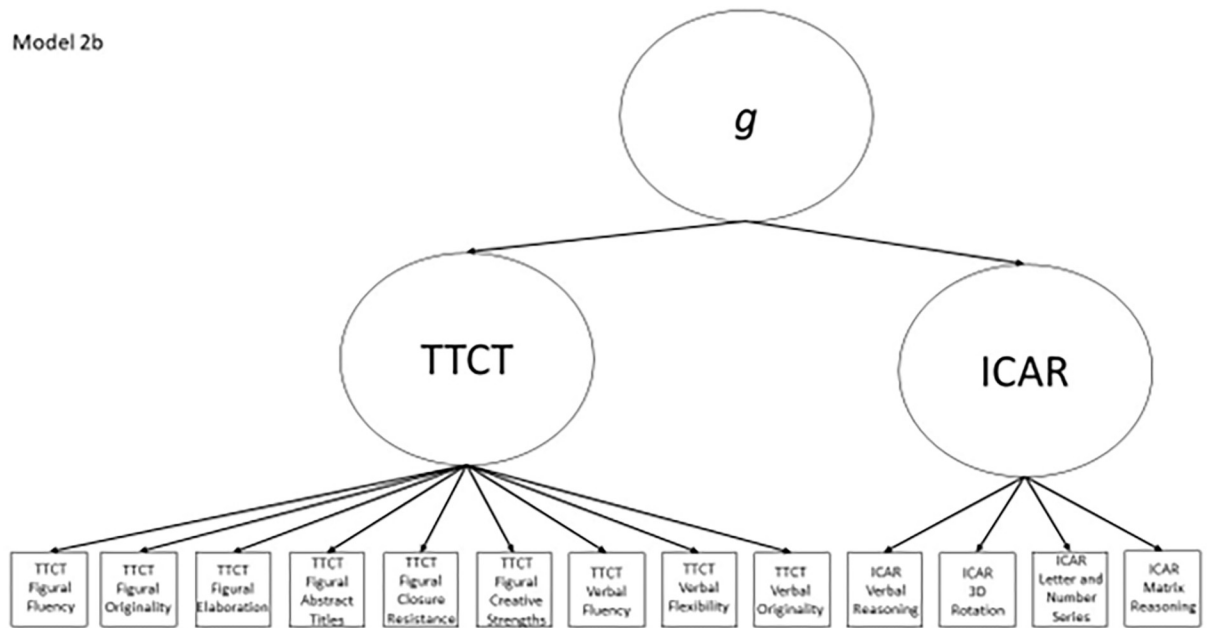


Fig 8. Confirmatory factor analysis diagram for model 2b.

<https://doi.org/10.1371/journal.pone.0274921.g008>

models, with Models 1a, 2a, 3a, and 4a nested within it. These latter models were to be compared to Model 5 in a sequential fashion by constraining a correlation between the relevant first-order factors to 1.0—in order to force these factors to merge—and conducting a chi-squared difference test with $k-1$ degree of freedom, where k is the number of factors being merged. There are two sequences of nested models that can be tested in this fashion. The first starts with the complex Model 4a, and testing whether it is statistically equal to Model 1a,

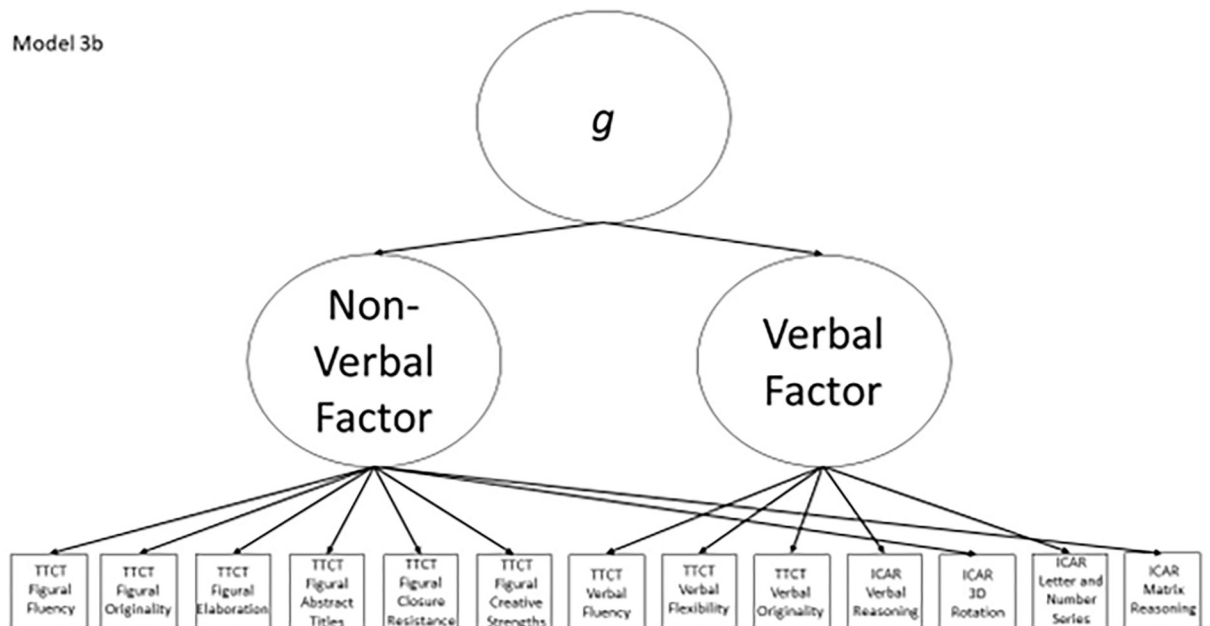


Fig 9. Confirmatory factor analysis diagram for model 3b.

<https://doi.org/10.1371/journal.pone.0274921.g009>

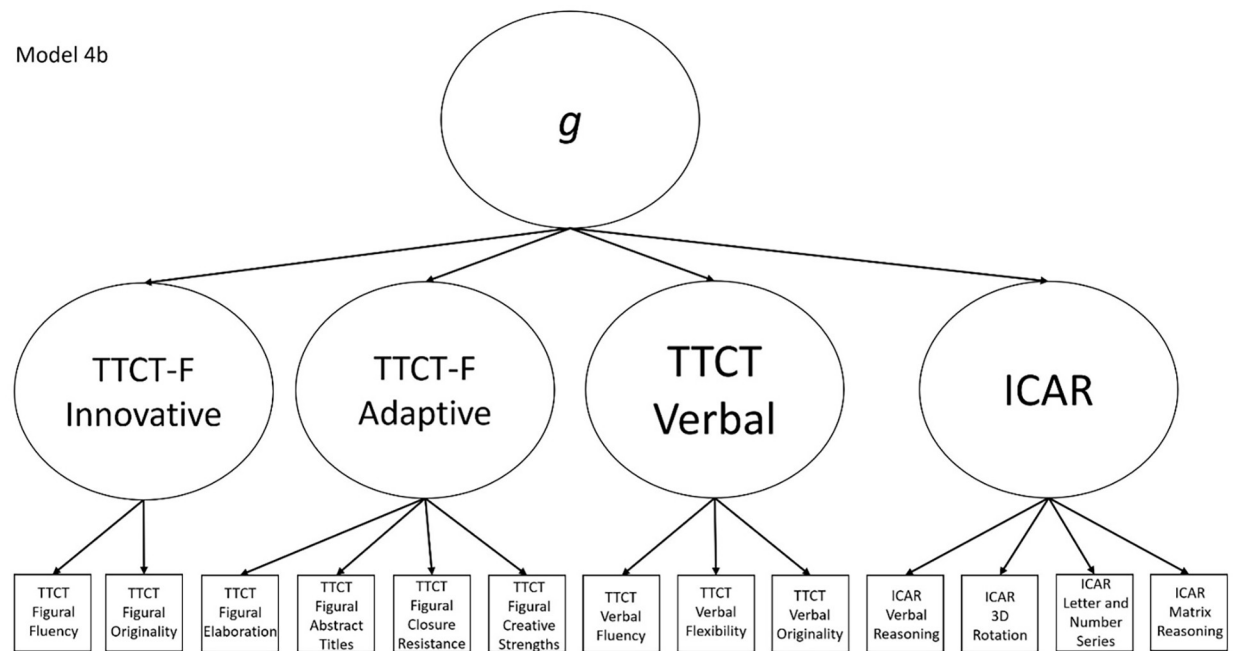


Fig 10. Confirmatory factor analysis diagram for model 4b.

<https://doi.org/10.1371/journal.pone.0274921.g010>

followed by a test to determine whether Model 2a is statistically equal to Model 5. The second group of nested models tested whether Model 3a is statistically equal to Model 5 in the same fashion. All of these comparisons were planned, on the condition that multiple nested models fit the data well (based on the fit statistics). Any tests that produce a non-statistically significant result (with $p > .05$) would indicate that the two nested models are equivalent and that the simpler model with more degrees of freedom should be preferred. Note that to test nested models, we constrained correlations between factors in the more complex model to be equal to 1.0 to create the more parsimonious model and test whether this model is statistically different from the more complex model. Because a constrained correlation of $r = 1.0$ creates a boundary constraint (because the correlation between factors cannot be greater than 1.0), we planned to use the appropriate mixed chi-squared distribution for each comparison (see [89] for details).

After identifying the best model(s) that fit the data, we interpreted the results in light of the different plausible relationships that divergent thinking test scores and intelligence test scores could have. If Models 1a or 2a fit best, that would indicate that the TTCT and ICAR measure separate but correlated constructs. If Model 3a fit best, then we would interpret this to indicate that the TTCT and ICAR measure a mix of verbal and non-verbal reasoning behaviors, whereas Model 4a being the best fitting model would support the view that TTCT tests measure different constructs and that the TTCT Figural test has a multifactorial structure. A best fitting Model 5 would indicate that all TTCT and ICAR tests are direct measures of g and that there are no separate constructs that these tests measure. If Models 1b or 2b fit the data best, then we would interpret this as indicating that the TTCT and ICAR measure separate first-order factors but that these then combine to form a second-order g factor. If Model 3b has the best fit indices, then this would mean that the tests measure a mix of verbal and non-verbal behaviors, which then coalesce into a g factor. Finally, if Model 4b fits best, then this would indicate that there is a general g factor among the subscores in the study, but that these subscores also form mediating factors for intelligence, verbal divergent thinking, an adaptive figural behavior, and innovative figural responses.

There was the possibility that no model will have adequate fit. If this occurs, we planned to interpret this result to indicate that the factor structure of divergent thinking and intelligence subtest scores does not conform to any *a priori* theorized structure and tests and fit statistic comparisons to find the best model are unneeded. The factor structure of these two tests in relation to each other would remain unresolved. In our registered report protocol, we intended not to engage in any exploratory analyses (e.g., exploratory factor analysis, modification of confirmatory models to improve fit) if no pre-specified confirmatory models were found to fit the data. As indicated later in this article, we later did conduct an exploratory analysis after obtaining approval from the article's editor.

Predictions

We had no firm predictions about what this study may show, and we were agnostic about the results. To us, all nine of these models were plausible, and we did not have a strong belief about which may be the best fitting model. We also recognized that because fit statistics are sensitive to different aspects of the data and of model misspecification that we could get contradictory indications of what “the best fitting model” is. For example, a model could have the lowest BIC but higher SRMR and RMSEA values than other models. If contradictory results appear in the fit statistics, we planned to report this and interpret it to indicate that we could not fully resolve the question of how divergent thinking test scores and intelligence subtest scores interrelate but that we had narrowed the range of possibilities. There were no attempts to modify models in order to improve fit.

Open science practices

This project is designed to conform to all standards of open science. All methods and materials (except those protected by copyright or confidentiality guidelines) are available to view at <https://osf.io/8rpfz/>. Publication as a Registered Report Protocol [80] functioned as a pre-registration for the study. To supplement final publication of the study, we have made all raw data, analysis syntax, and output files freely available online at <https://osf.io/8rpfz/>. We have also made a pre-print of the results available when submitting to the journal. Additionally, we will provide the response forms available to researchers who explain their qualifications and interests to us, so that they can re-analyze our data with alternate scoring methods (such as those explained by Reiter-Palmon et al. [43]). Readers should note that making response booklets and examinee responses openly available online would violate copyright on the TTCT and item confidentiality of both tests, so we will control access to these materials.

Results

Descriptive statistics

Table 2 lists the means, standard deviations, and the reliability coefficients (i.e., Cronbach's α values) on all analyzed variables. Compared to the averages for the norm sample, the mean raw subscores of our sample were higher than average for the TTCT-V ($d = .18$ to $.50$) and lower than average for the TTCT-F subscores ($d = -.74$ to $-.05$). These effect sizes were calculated from raw scores that had been converted to standard scores that have a mean of 100 and standard deviation of 20 for each subscore. A summary report of mean standard scores for the sample is available on the study's OSF page at <https://osf.io/8rpfz/>. Table 3 is a correlation table of the same variables for our sample members. For both of these tables, $n = 427$ sample members who had complete data on all variables in the table.

Table 2. Descriptive statistics of variables used in the CFA models.

Variable	Min	Max	M	SD	α
TTCT-F Fluency (raw score)	3.00	40.00	18.73	6.57	—
TTCT-F Originality (adjusted score)	1.50	11.78	2.64	1.47	0.732
TTCT-F Elaboration (adjusted score)	1.24	6.33	1.75	0.59	0.707
TTCT-F Abstract Titles (adjusted score)	0.00	2.00	1.25	0.46	—
TTCT-F Closure (adjusted score)	0.00	2.00	1.70	0.24	—
TTCT-F Creative Strengths Checklist (raw score)	2.00	24.00	14.80	3.14	0.715
TTCT-V Fluency (raw score)	12.00	183.00	88.49	29.73	0.854
TTCT-V Originality (adjusted score)	3.01	6.00	4.57	0.50	0.843
TTCT-V Flexibility (adjusted score)	1.96	5.46	3.22	0.48	0.732
ICAR Verbal Reasoning (raw score)	1.00	15.00	8.61	2.48	0.636
ICAR 3D Rotation (raw score)	0.00	23.00	3.81	4.63	0.891
ICAR Letter-Number Series (raw score)	0.00	9.00	4.26	2.45	0.756
ICAR Matrix Reasoning (raw score)	0.00	10.00	4.99	2.36	0.622

<https://doi.org/10.1371/journal.pone.0274921.t002>

Table 4 displays the fit statistics for Models 1–5. Among the six models which had converging results, Models 1a, 1b, and 4a had the best values for their fit indices. However, all of these models had specification problems. Models 1a and 1b had a negative residual variance for the TTCT-F adjusted elaboration score. Model 1b also had the *g* factor defined by higher-order loadings that were all statistically equal to zero (all two-tailed *p*s $\geq .939$). For Model 4a, the factor loadings (and correlation coefficients with other latent factors) for the TTCT-F Adaptive factor were undefined.

Moreover, none of the models met the pre-registered standards for acceptable model fit (i.e., an SRMR value $\leq .08$ and at least one of the following: (1) CFI $\geq .90$, (2) TLI $\geq .90$, or (3) RMSEA $\leq .08$). By these standards, none of the models can be defined as representing the data well. The S1 File reports results of the same models with the creative strengths checklist score. These models also did not reach the standards for sufficient model fit.

Table 3. Correlation table for variables used in the CFA models.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. TTCT-F Fluency (raw score)	1.00												
2. TTCT-F Originality (adjusted score)	-.315	1.00											
3. TTCT-F Elaboration (adjusted score)	-.121	.391	1.00										
4. TTCT-F Abstract Titles (adjusted score)	-.015	.020	.254	1.00									
5. TTCT-F Closure (adjusted score)	.007	-.052	.039	.067	1.00								
6. TTCT-F Creative Strengths Checklist (raw score)	.215	-.026	.370	.394	.163	1.00							
7. TTCT-V Fluency (raw score)	.448	-.001	.173	.077	-.026	.349	1.00						
8. TTCT-V Originality (adjusted score)	.132	.064	.044	.038	-.097	.076	.292	1.00					
9. TTCT-V Flexibility (adjusted score)	-.354	-.020	-.083	-.085	-.028	-.225	-.706	-.256	1.00				
10. ICAR Verbal Reasoning (raw score)	-.017	-.014	.070	.097	.077	.094	.046	.011	-.071	1.00			
11. ICAR 3D Rotation (raw score)	.020	-.044	.007	.074	.016	.117	.003	-.053	-.005	<u>.307</u>	1.00		
12. ICAR Letter-Number Series (raw score)	-.005	-.092	.023	.080	.051	.097	-.002	-.026	.004	<u>.502</u>	<u>.389</u>	1.00	
13. ICAR Matrix Reasoning (raw score)	-.010	-.073	-.007	.083	.109	.152	.042	-.029	-.019	<u>.333</u>	<u>.316</u>	<u>.407</u>	1.00

Note. Intercorrelations of scores from the same test are given the same emphasis (i.e., bold, italics, or underline).

Note. *n* = 427; *p* < .05 for correlations $r \leq \pm.095$; *p* < .01 for correlations $r \geq \pm.125$; *p* < .001 for correlations $r \geq \pm.159$ (all *p*-values two-tailed).

<https://doi.org/10.1371/journal.pone.0274921.t003>

Table 4. Fit statistics for CFA models.

Model	χ^2	df	CFI	TLI	RMSEA [90% CI]	SRMR	AIC	BIC
1a	326.608	62	.809	.759	.100 [.089, .111]	.096	20,522.117	20,692.502
1b	326.610	62	.809	.759	.100 [.089, .111]	.096	20,522.119	20,692.504
2a	1,459.916	78	.568	.473	.148 [.138, .158]	.107	20,852.709	21,014.980
2b	—	—	—	—	—	—	—	—
3a	1,459.916	78	.593	.505	.143 [.133, .154]	.132	20,817.338	20,979.609
3b	—	—	—	—	—	—	—	—
4a	1,459.916	78	.827	.771	.097 [.087, .108]	.090	20,499.376	20,681.931
4b	—	—	—	—	—	—	—	—
5	1,459.916	78	.224	.083	.195 [.185, .205]	.149	21,325.971	21,480.129

Note. Models 2b, 3b, and 4b did not converge.

<https://doi.org/10.1371/journal.pone.0274921.t004>

Discussion

We undertook this study in an attempt to discern whether the TTCT tests and a cognitive ability test—in this case, the ICAR—measure the same construct or different constructs. To do so, we created a series of confirmatory factor analysis models (shown in Figs 2–10) that could test a variety of theoretical perspectives about the possible relationships among subscores on divergent thinking and intelligence tests. These perspectives reflected:

- the theory that creativity and intelligence were separate constructs that could be measured coherently with different psychometric tests (Models 1a and 2a),
- the proposed split between verbal and non-verbal abilities (Model 3a),
- a view that the TTCT-F had groups of subscores that measured innovative and adaptive components that were separate from other abilities in the domains of creativity and intelligence (Model 4a),
- variations of the above models that included a general intellectual ability at the top of a hierarchy (Models 1b, 2b, 3b, and 4b), and
- a simple view that all subtests in the study's divergent thinking and intelligence were direct measures of a global intellectual ability (Model 5).

To strengthen confidence in the study and reduce subjectivity in the methods, we pre-registered the study's methodology by publishing the study as a registered report [80]. The proposed methodology was carried out successfully without complications.

The results of our study are inconclusive because none of the nine plausible models met the *a priori* standards of good model fit. Recognizing this, we tentatively suggest that Model 1a—in which the ICAR, TTCT-F, and TTCT-V measure different underlying latent variables—comes closest to approximating the data, though it has model fit problems. This provides some evidence that the tests may measure different constructs, though the poor model fit prevents us from confidently committing ourselves to this viewpoint. Thus, the study failed to answer the question we had, and the question of whether tests of divergent thinking—in this case the TTCT-V and TTCT-F—measure intelligence remains unresolved.

To understand the factor structure of the test scores better, we conducted an unplanned post hoc confirmatory factor analysis of a congeneric model for each test separately (with the editor's permission). We performed this exploratory analysis in order to determine whether the lack of model fit in the pre-specified models was due to (1) measurement problems with

one or more of the constituent tests, or (2) misspecification of the interrelationships of observed variables in the pre-specified models. The TTCT-F congeneric measurement model showed poor fit ($\chi^2 = 608.540$, $df = 15$, CFI = .826, TLI = .710, RMSEA = .164 [90% CI = .138, .192], SRMR = .093, AIC = 7,380.418, BIC = 7,453.441). Incorporating this TTCT-F factor into a larger model will greatly reduce overall model fit. Therefore, our study is inconclusive because the TTCT-F does not seem to measure *any* unitary coherent construct, meaning the question of whether it also measures intelligence is moot. From a theoretical perspective, the lack of model fit in the pre-specified models is due (at least partially) to measurement problems in the TTCT-F. Practically speaking, our results raise serious questions about what the TTCT-F measures. Because TTCT-F scores fail to conform to a coherent unitary measurement model, the test likely does not measure a single construct (e.g., creativity or divergent thinking). Indeed, our study throws serious doubt on the internal validity of TTCT-F scores. Composite creativity index scores or overall average scores produced by the TTCT-F (not used in this study, but sometimes used in educational and research settings) are probably uninterpretable.

The TTCT-V congeneric measurement model was saturated with zero degrees of freedom, making it uninformative by itself. However, *post hoc* (i.e., not pre-registered) examination of the correlations among TTCT-V variables in Table 3 shows that the fluency and flexibility scores are strongly correlated ($r = -.706$), but that both are rather weakly correlated with originality ($r = .292$ and $-.256$ for fluency and flexibility, respectively). The fact that TTCT-V fluency scores are negatively correlated with the other two seems to indicate that there is a trade-off, where generating a large number of responses tends to reduce scores on the flexibility and originality aspects of divergent thinking. This raises the possibility that the TTCT-V may also not measure a coherent, unitary construct, though the saturated congeneric measurement model means we cannot investigate this further with our data.

In contrast, the ICAR showed excellent model fit ($\chi^2 = 2.093$, $df = 2$, CFI = 1.00, TLI = 0.99, RMSEA = .010 [90% CI = .000, .097], SRMR = .013, AIC = 8,146.985, BIC = 8,195.667). This test clearly does measure a single latent construct—general intelligence. Therefore, the poor model fits shown in Table 4 are not due to measurement problems arising from the ICAR. The ICAR can confidently be said to measure a single construct, and overall ICAR scores are good proxies of a respondent's level of intelligence.

Limitations

One advantage of our study is also a limitation: as authors of a registered report, we were locked into a scoring system and analysis plan. This strengthened our confidence in our findings because we knew they could not be the product of questionable research practices or unconscious adjustments to our study as we analyzed the data. However, some of the analysis decisions are still arbitrary, such as the method used to adjust scores for the confounding influence of fluency (which greatly reduced the magnitude of correlations among TTCT variables and flipped many from positive to negative correlations). Because there are other scoring alterations which can control for fluency, we have made our altered scores and the original scores available in our open data set at <https://osf.io/8rpfz/> so that researchers can use other methods for controlling for the confound of fluency and test our models (or other hypotheses) accordingly. Alternate methods of controlling for fluency may be useful for identifying causes of model misfit in our study and improve the interpretation of TTCT scores.

Prior studies using raw scores—instead of scores adjusted for the influence of fluency—have shown that TTCT-F data can fit confirmatory factor analysis models [47, 48]. The differing results in our study may indicate that a well-fitting measurement model with raw TTCT

scores may be an artifact of the confounding influence of the fluency score on the other scores. If this is true, then it raises fundamental questions about the validity of interpretations of originality, flexibility, and other non-fluency TTCT scores.

This study is also limited by the convenience sample. Most participants were college students, though the open enrollment nature of the university makes restriction of range much less of a problem for the intelligence test scores than at many universities. Yet, any restriction of range did not prevent the ICAR data from having an extremely good model fit by itself. We have no reason to suspect that the convenience sample suffered from restriction of range in the creativity scores. Score summaries (available at <https://osf.io/8rpfz/>) show that standard scores spanned a very wide range of scores and that there were no apparent ceiling or floor effects. Yet, the degree of generalizability of these findings is still unclear.

Finally, it is important to recognize that there is some subjectivity inherent with scoring some of the tasks on the TTCT. For example, the TTCT-F's creative strengths checklist requires scorers to make judgements about the humor, richness, expressiveness of titles, and other aspects of responses. Even core aspects of the TTCT-F and TTCT-V, such as originality, can require subjective judgment in the scoring process, especially when examinees' responses are somewhat ambiguous. We believe that using professional scorers employed by the testing company—who have extensive training and experience with the TTCT—would minimize subjectivity in the scoring process and increase the consistency of subjective judgements. It is also important to note that subjectivity in scoring is not a unique problem with the TTCT and that other measures of creativity and divergent thinking have the same problem. Researchers cannot eliminate subjectivity in scoring creativity tests; the best option is to manage and minimize it, and we have tried to do so as much as possible.

Conclusion

Although our results were unable to answer the original research question of whether creativity and intelligence tests measure the same construct, we do believe that our results raise serious questions about the construct validity evidence of the TTCT-F and, to a lesser extent, the TTCT-V. This finding adds to earlier evidence [90] showing that variance of TTCT scores may be strongly influenced by item format, content, and demand characteristics of specific TTCT tasks. Additionally, the poor fit of each TTCT test's subscores to a congeneric measurement model seriously undermines the support for interpreting and using a global divergent thinking score. Any apparent, strong positive correlations among raw scores on the TTCT are likely a consequence of the confounding effect of fluency scores on the other scores. Our research presents a serious challenge to the validity of interpreting TTCT-V and TTCT-F scores as measuring a coherent latent construct. We recommend against using overall TTCT-F scores for non-research purposes and believe that creativity scholars should be skeptical about the ability of the TTCT-F to measure a coherent underlying construct. We also recommend caution in using TTCT-V scores for making decisions about examinees. Until there is a coherent measurement model, a scoring system, and interpretation that aligns with that model and with theory, the TTCT-F (and possibly the TTCT-V) should be confined to research settings only.

More research is needed to determine whether divergent thinking tests and other operationalizations of creativity function as measures of intelligence. For the TTCT specifically, the psychometric properties of the tests need to be improved (especially for the TTCT-F). The quality of any research cannot surpass the quality of the data; when instruments fall short of psychometric standards, the scores can reveal little about people's thought processes. Once the TTCT has better measurement properties, psychologists will be able to determine whether it functions as a measurement of intelligence (or any other construct, for that matter). Future

researchers can also apply our methods to other measures of creativity and/or divergent thinking to determine whether they measure intelligence.

Supporting information

S1 File. Confirmatory factor analysis results for models that include the TTCT-F creative strengths checklist.
(PDF)

Acknowledgments

We appreciate the efforts of Andrew Rosenbaum and Elizabeth Bradley in administering TTCT tests.

Author Contributions

Conceptualization: Russell T. Warne.

Data curation: Russell T. Warne.

Formal analysis: Russell T. Warne.

Funding acquisition: Russell T. Warne, Sam Golightly.

Investigation: Russell T. Warne, Makai Black.

Methodology: Russell T. Warne.

Project administration: Russell T. Warne, Sam Golightly, Makai Black.

Resources: Russell T. Warne.

Supervision: Russell T. Warne.

Validation: Russell T. Warne.

Writing – original draft: Russell T. Warne, Sam Golightly.

Writing – review & editing: Russell T. Warne, Sam Golightly, Makai Black.

References

1. Guilford JP. Creativity. *Am Psychol.* 1950; 5: 444–54.
2. Kell HJ, Lubinski D, Benbow CP, Steiger JH. Creativity and technical innovation: Spatial ability's unique role. *Psychol Sci.* 2013; 24: 1831–6. <https://doi.org/10.1177/0956797613478615> PMID: 23846718
3. Sternberg RJ, Grigorenko EL. Guilford's structure of intellect model and model of creativity: Contributions and limitations. *Creat Res J.* 2001; 13: 302–16.
4. Guilford JP. Some changes in the structure-of-intellect model. *Educ Psychol Meas.* 1988; 48: 1–4.
5. Silvia P. Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan (1965). *Creat Res J.* 2008; 20:34–39.
6. Karwowski M, Dul J, Gralewski J, Jauk E, Jankowska DM, Gajda A, et al. Is creativity without intelligence possible? A necessary condition analysis. *Intelligence.* 2016; 57: 105–117.
7. Meier MA, Burgstaller JA, Benedek M, Vogel SE, Grabner RH. *J Intell.* 2021; 9: 10.
8. Weiss S, Steger D, Schroeders U, Wilhelm O. *J Intell.* 2020; 8: 38.
9. Silvia PJ. Intelligence and creativity are pretty similar after all. *Educ. Psych. Rev.* 2015; 27: 599–606.
10. Wallach MA, Kogan N. *Modes of thinking in young children: A study of the creativity–intelligence distinction.* New York: Holt, Rinehart, & Winston.
11. Plucker JA. Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data. *Creat Res J.* 1999; 12:103–14.

12. Torrance EP. Career patterns and peak creative achievements of creative high school students twelve years later. *Gift Child Q.* 19;75–88.
13. Weiss S, Steger D, Kaur Y, Hildebrandt A, Schroeders U, Wilhelm O. *Eur J Pers.* Forthcoming 2021.
14. Preckel F, Wermer C, Spinath FM. The interrelationship between speeded and unspeeded divergent thinking and reasoning, and the role of mental speed. *Intelligence.* 2011; 39: 378–388.
15. Lee CS, Therriault DJ. The cognitive underpinnings of creative thought: A latent variable analysis exploring the roles of intelligence and working memory in three creative thinking processes. *Intelligence.* 2013; 41: 306–320.
16. Süß H-M, Oberauer K, Wittmann WW, Wilhelm O, Schulze R. Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence.* 2002; 30: 261–288.
17. Sowden PT, Pringle A, Gabora L. The shifting sands of creative thinking: Connections to dual-process theory. *Think Reason.* 2015; 21:40–60.
18. Simonton DK. *Creativity: Cognitive, personal, developmental, and social aspects.* 2000; 55:151–8.
19. Galton F. *Inquiries into human faculty and its development.* London: MacMillan; 1883.
20. Runco M, Jaeger G. The standard definition of creativity. *Creat Res J.* 2012; 24:92–6.
21. Stein MI. Creativity and culture. *J Psychol.* 1953; 36:311–22.
22. Runco MA. Commentary: Divergent thinking is not synonymous with creativity. *Psychol Aesthet Creat Arts.* 2008; 2:93–96.
23. McCrae RR. Creativity, divergent thinking, and openness to experience. *J Pers Soc Psychol.* 1987; 52:1258–65.
24. Vartanian O, Martindale C, Matthews J. Divergent thinking ability is related to faster relatedness judgments. *Psychol Aesthet Creat Arts.* 2009; 3:99–103.
25. Glăveanu VP. Creativity as cultural participation. *J Theory Soc Behav.* 2011; 41:48–67.
26. Primi R, Silvia PJ, Jauk E., Benedek M Applying many-facet Rasch modeling in the assessment of creativity. *Psychol Aesthet Creat Arts.* 2019; 13:176–86.
27. Runco M, Millar G, Acar S, Cramond B. Torrance Tests of Creative Thinking as predictors of personal and public achievement: A fifty-year follow-up. *Creat Res J.* 2010; 22:361–8.
28. Corazza GE. Potential originality and effectiveness: The dynamic definition of creativity. *Creat Res J.* 2016; 28:258–267.
29. Furnham A, Bachtiar V. Personality and intelligence as predictors of creativity. 2008; 45:613–7.
30. Furnham A, Niderstrom M. Ability, demographic and personality predictors of creativity. *Pers Individ Differ.* 2010; 48:957–61.
31. Wang MZ, Chen W, Zhang C, Deng XL. Personality types and scholarly creativity in undergraduate students: The mediating roles of creative styles. *Pers Individ Differ.* 2017; 105:170–4.
32. Benedek M, Borovnjak B, Neubauer AC, Kruse-Weber S. Creativity and personality in classical, jazz and folk musicians. *Pers Individ Differ.* 2014; 63:117–121. <https://doi.org/10.1016/j.paid.2014.01.064> PMID: 24895472
33. Fink A, Woschnjak S. Creativity and personality in professional dancers. *Pers Individ Differ.* 2011; 51:754–8.
34. Mammadov S, Cross TL, Cross JR. In search of temperament and personality predictors of creativity: A test of a mediation model. *Creat Res J.* 2019; 31:174–87.
35. Nečka E., Hlawacz T. Who has an artistic temperament? Relationships between creativity and temperament among artists and bank officers. *Creat Res J.* 2013; 25:182–8.
36. Hannam K, Narayan A. Intrinsic motivation, organizational Justice, and creativity. *Creat Res J.* 2015; 27:214–24.
37. Saether EA. Creativity-contingent rewards, intrinsic motivation, and creativity: The importance of fair reward evaluation procedures. *Frontiers in Psychology.* 2020; 11:974. <https://doi.org/10.3389/fpsyg.2020.00974> PMID: 32595550
38. Silvia PJ, Christensen AP, Cotter KN. Commentary: The development of creativity-ability, motivation, and potential. *New Dir Child Adolesc Dev.* 2016; 151:111–9.
39. Mumford MD, Gustafson SB. Creativity syndrome: integration, application, and innovation. *Psychol Bull.* 1988; 103:27–43.
40. Vygotsky LS. Imagination and creativity in childhood. *J Russ East Eur Psychol.* 2004; 42:1–92.
41. Sawyer RK. *Explaining creativity: The science of human innovation.* 2nd ed. New York: Oxford University Press.

42. Davies D, Jindal-Snape D, Collier C, Digby R, Hay P, Howe A. Creative learning environments in education—A systematic literature review. *Think Skills Creat*. 2013; 8:80–91.
43. Reiter-Palmon R, Forthmann B, Barbot B. Scoring divergent thinking tests: A review and systematic framework. *Psychol. Aesthet. Creat. Arts*. 2019; 13: 144–152.
44. Silvia PJ, Winterstein BP, Willse JT. Rejoinder: the madness to our method: Some thoughts on divergent thinking. *Psychol Aesth Creat Arts*. 2008; 2:109–14.
45. Clapham MM. The convergent validity of the Torrance Tests of Creative Thinking and Creativity Interest Inventories. *Educ Psychol Meas*. 2004; 64:828–41.
46. Bart WM, Hokanson B., Can I. An investigation of the factor structure of the Torrance Tests of Creative Thinking. *Educ Sci Theory Pract*. 2017; 17:515–28.
47. Humble S, Dixon P, Mpofu E. Factor structure of the Torrance Tests of Creative Thinking Figural Form A in Kiswahili speaking children: Multidimensionality and influences on creative behavior. *Think Skills Creat*. 2018; 27:33–44.
48. Kim KH. Is creativity unidimensional or multidimensional? Analyses of the Torrance Tests of Creative Thinking. *Creat Res J*. 2006; 18:251–9.
49. Kim KH, Cramond B, Bandalos DL. The latent structure and measurement invariance of scores on the Torrance Tests of Creative Thinking—Figural. *Educ Psychol Meas*. 2006; 66:459–77.
50. Said-Metwaly S, Fernández-Castilla B, Kyndt E, Van den Noortgate W. The factor structure of the Figural Torrance Tests of Creative Thinking: A meta-confirmatory factor analysis. *Creat Res J*. 2018; 30:352–60.
51. Kirton M. Adaptors and innovators: A description and measure. *J Appl Psychol*. 1976; 61:622–9.
52. Sternberg RJ, Detterman DK, editors. *What is intelligence?* Norwood, NJ: Ablex Publishing; 1986.
53. Gottfredson LS. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*. 1997; 24: 13–23.
54. Perkins DN, Farady M, Bushey B. Everyday reasoning and the roots of intelligence. In: Voss JF, Perkins DN, Segal JW, editors. *Informal reasoning and education*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1991. pp. 83–105.
55. Rowe DC. A place at the policy table? Behavior genetics and estimates of family environmental effects on IQ. *Intelligence*. 1997; 24: 133–158.
56. Jensen AR. The theory of intelligence and its measurement. *Intelligence*. 2011; 39: 171–177.
57. Mpofu E. Being intelligent with Zimbabweans. In: Sternberg RJ, editor. *New York, NY: Cambridge University Press; 2004. pp. 364–390.*
58. Scarr S. Intelligence: Revisited. In: Sternberg RJ, Detterman DK, editors. *Norwood, NJ: Ablex Publishing; 1986. pp. 117–120.*
59. Floyd RG, Reynolds MR, Farmer RL, Kranzler JH. Are the general factors from different child and adolescent intelligence tests the same? Results from a five-sample, six-test analysis. *Sch. Psychol. Rev*. 2013; 42: 383–401.
60. Johnson W, Bouchard TJ, Krueger RF, McGue M, Gottesman II. Just one *g*: Consistent results from three test batteries. *Intelligence*. 2004; 32: 95–107.
61. Johnson W, te Nijenhuis J, Bouchard TJ. Still just 1 *g*: Consistent results from five test batteries. *Intelligence*. 2008; 36: 81–95.
62. Keith TZ, Kranzler JH, Flanagan DP. What does the Cognitive Assessment System (CAS) measure? Joint confirmatory factor analysis of the CAS and the Woodcock-Johnson Tests of Cognitive Ability (3rd edition). *Sch. Psychol. Rev*. 2001; 30: 89–119.
63. Stauffer JM, Ree MJ, Carretta TR. Cognitive-components tests are not much more than *g*: An extension of Kyllonen's analyses. *J. Gen. Psychol*. 1996; 123: 193–205.
64. Warne RT. *In the know: Debunking 35 myths about human intelligence*. New York: Cambridge University Press; 2020.
65. Warne RT, Burningham C. Spearman's *g* found in 31 non-Western nations: Strong evidence that *g* is a universal phenomenon. *Psychol Bull*. 2019; 145:237–272. <https://doi.org/10.1037/bul0000184> PMID: 30640496
66. Carroll JB. *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press; 1993.
67. Spearman C. "General intelligence," objectively determined and measured. *Am J Psychol*. 1904; 15:201–93.
68. Spearman C. *The abilities of man: Their nature and measurement*. New York: Macmillan Company; 1927.

69. Gordon RA. Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence*. 1997; 24:203–320.
70. Gottfredson LS. Why *g* matters: The complexity of everyday life. *Intelligence*. 1997; 24:79–132.
71. Lubinski D. Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "General intelligence, objectively determined and measured." *J Pers Soc Psychol*. 2004; 86:96–111.
72. Sanders CE, Lubinski D, Benbow CP. Does the Defining Issues Test measure psychological phenomena distinct from verbal ability? An examination of Lykken's query. *J Pers Soc Psychol*. 1995; 69:498–504.
73. Gottfredson LS. Intelligence: Is it epidemiologists' elusive "fundamental cause" of social class inequalities in health? *J Pers Soc Psychol*. 2004; 86:174–99.
74. Sternberg RJ, O'Hara LA. Creativity and intelligence. In: Sternberg RJ, editor. New York, NY: Cambridge University Press; 1999. pp. 251–272.
75. Corazza GE, Lubart T. Intelligence and creativity: Mapping constructs on the space-time continuum. *J. Intell*. 2021; 9: 1.
76. Binet A, Simon T. Le développement de l'intelligence chez les enfants. *L'Année psychologique*. 1911; 17: 145–201. Reprinted In: Binet A, Simon T. The development of intelligence in the child. Kite ES, translator. Baltimore, MD: Williams & Wilkins Company. 1916; pp. 182–273.
77. Torrance Tests of Creative Thinking streamlined scoring guide for figural forms A and B. Bensenville, IL: Scholastic Testing Service; 2017.
78. Condon DM, Revelle W. The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*. 2014; 43:52–64.
79. Gibbons A, Warne RT. First publication of subtests in the Stanford-Binet 5, WAIS-IV, WISC-V, and WPPSI-IV. *Intelligence*. 2019; 75:9–18.
80. Warne RT, Golightly S, Black M. Factor structure of intelligence and divergent thinking subtests: A registered report. *PLoS ONE*. 2021; 16(5):0251268. <https://doi.org/10.1371/journal.pone.0251268> PMID: 33979390
81. Forthmann B, Jendryczko D, Scharfen J, Kleinkorres R, Benedek M, Holling H. Creative ideation, broad retrieval ability, and processing speed. *Intelligence*. 2019; 75:59–72.
82. Kline RB. Principles and practice of structural equation modeling. New York: Guilford Press; 2005.
83. Hancock GR. Power analysis in covariance structure modeling. In: Hancock GR, Mueller RO, editors. Structural equation modeling: A second course. Greenwich, CT; 2006. p. 69–115.
84. Kiyonaga A, Scimeca JM. Practical considerations for navigating registered reports. *Trends Neurosci*. 2019; 42:568–72. <https://doi.org/10.1016/j.tins.2019.07.003> PMID: 31470913
85. Kim KH. Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creat. Res. J*. 2006; 18: 3–14.
86. Jensen AR. The *g* factor: The science of mental ability. Westport, CT: Praeger; 1998.
87. Fan X, Sivo SA. Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Struct Equ Modeling*. 2005; 12:343–67.
88. Sun J. Assessing goodness of fit in confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development*. 2005; 37:240–56.
89. Stoel RD, Garre FG, Dolan C, van den Wittenboer G. On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychol. Methods*. 2006; 11: 439–455. <https://doi.org/10.1037/1082-989X.11.4.439> PMID: 17154756
90. Almeida LS, Prieto LP, Ferrando M, Oliveira E, Ferrándiz C. Torrance Test of Creative Thinking: The question of its construct validity. *Think Skills Creat*. 2008; 3:53–8.