Empirical Paper

On the trail of creativity: dimensionality of divergent thinking and its relation with cognitive abilities, personality, and insight

S. Weiss[1], D. Steger[1], Y. Kaur[2], A. Hildebrandt[2], U. Schroeders[3] and O. Wilhelm[1]

Abstract

Divergent thinking (DT) is an important constituent of creativity that captures aspects of fluency and originality. The literature lacks multivariate studies that report relationships between DT and its aspects with relevant covariates, such as cognitive abilities, personality traits (e.g. openness), and insight. In two multivariate studies (N = 152 and N = 298), we evaluate competing measurement models for a variety of DT tests and examine the relationship between DT and established cognitive abilities, personality traits, and insight. A nested factor model with a general DT and a nested originality factor described the data well. In Study 1, DT was moderately related with working memory, fluid intelligence, crystallized intelligence, and mental speed. In Study 2, we replicate these results and add insight, openness, extraversion, and honesty–humility as covariates. DT was associated with insight, extraversion, and honesty–humility, whereas crystallized intelligence mediated the relationship between openness and DT. In contrast, the nested originality factor (i.e. the specificity of originality tasks beyond other DT tasks) had low variance and was not meaningfully related with any other constructs in the nomological net. We highlight avenues for future research by discussing issues of measurement and scoring.

Introduction

For over a century, researchers are trying to assess and understand creativity (e.g. Patrick, 1935), which has been related to both typical behaviour (e.g. personality; Guilford, 1950) and maximal effort (e.g. intellect; Guilford, 1967). In the last years, the importance of creativity has been stressed with respect to several crucial outcomes, from academic achievement (Gajda, Karwowski, & Beghetto, 2017) to affective disorders (Acar & Sen, 2013; Taylor, 2017). Moreover, creativity has been also described as a crucial human source of action in work context (PWC, 2016): hence, an increasing number of studies are examining it within a school context. For example, creative thinking assessment has been included in the innovative domain for the upcoming PISA 2021 study (see ACT, n.d.; Barbot, Hass, & Reiter-Palmon, 2019). Despite its growing societal relevance, creativity remains poorly understood as a construct, even after over half a century of research. Hence, we aim

to better understand creativity and ways to assess it. One way to do so is to embed creativity in the nomological net of established abilities and traits. The purpose of the present studies is to improve our understanding of creativity as a unique construct and individual differences in creativity.

Although studied for over a century, there is surprisingly little consensus regarding the measurement and scoring of creativity and its relation with other established constructs—such as cognitive abilities (Benedek, Könen, & Neubauer, 2012; Forthmann,

---

[1] Institute of Psychology and Education, Ulm University, Ulm, Germany
[2] Department of Psychology, Carl von Ossietzky University Oldenburg, Oldenburg, Germany
[3] Institute of Psychology, University of Kassel, Kassel, Germany

Corresponding author:
*Selina Weiss, Institute for Psychology and Pedagogy, Ulm University, Albert-Einstein Strasse 47, 89081 Ulm, Germany.
Email: selina.weiss@uni-ulm.de

Holling, Çelik, Storme, & Lubart, 2017; Jäger, Süß, & Beauducel, 1997; Silvia, Beaty, & Nusbaum, 2013; Süß & Beauducel, 2005) and personality traits (Barron & Harrington, 1981; Batey & Furnham, 2006; Feist, 1998; Guilford, 1950; McCrae, 1987). Divergent thinking (DT) tasks have been widely applied as measures of creativity. Fluency and originality have often been proposed as core aspects of DT (Carroll, 1993). The internal structure of DT tasks and their relations with other abilities and traits in the nomological net are subject of an ongoing debate (Silvia et al., 2013).

With our paper, we address two research questions: first, can originality (the quality of ideas indicated by their rareness, novelty, and unusualness) be distinguished from fluency? Second, to what extent is DT (based on indicators of fluency and originality) related with established cognitive abilities, personality traits, and insight? To answer these questions, we establish and compare competing measurement models of DT. To further our understanding, we then juxtapose DT with cognitive abilities and personality traits. Additionally, insight is added to this nomological net, as insight has conceptual similarities with creativity and intelligence. Taken together, this article contributes to the debate on the dimensionality and validity of creative abilities.

What is creativity?

The scientific study of creativity in psychology was (re-)-popularized after Guilford's (1950) presidential address to the members of the *American Psychological Association*. Since then, different branches of creativity research within psychology have developed, all of which accompanied by numerous definitions. Mostly, creativity is defined as a product or an idea that is original and therefore new, unusual, novel, or unexpected, and that is deemed valuable, useful, or appropriate (e.g. Barron, 1955; Batey, 2012; Mumford, 2003; Runco & Jaeger, 2012; Stein, 1953). A commonality of many branches in creativity research is that the generation of novel ideas is seen as pivotal for creative ability (Runco & Jaeger, 2012). Another consensual aspect of this concept of creativity is that, besides novelty, ideas are deemed creative if they are statistically infrequent, rare, or unexpected. A further consensual aspect is the usefulness and appropriateness of a creative product. Originality is therefore not only necessary but it is also a key characteristic when determining the degree of creativity (Abraham, 2018). Taken together, originality is a central and broadly accepted element of creativity, but originality might not exhaust all aspects of creativity (Abraham, 2018).

Individual differences in creativity can be seen in the processes (Barbot, 2018; Simonton, 2011)[1] and in the products of highly creative persons (Amabile, 1982), as well as in the creative ability of a person (e.g. creative test performance, Kandler et al., 2016).

In the present studies, we focus on the creative ability of persons, which is also referred to as a person's creative potential (Sternberg & Lubart, 1993). We stick to the terminology of ability as the measurement of DT that provides an assessment of such. Creative ability, as measured by DT tasks, requires generating specific ideas to solve a given problem (Guilford, 1967). DT is an essential constituent of creativity that entails the generation of original and novel ideas and products (Guilford, 1950, 1966; Lubart, 2001; Lubart, Pacteau, Jacquet, & Caroff, 2010; Runco & Acar, 2012).

How is divergent thinking structured?

DT tasks have been widely applied as an assessment substitute of real-world creativity (Runco & Acar, 2012). They were designed to capture the fluency, flexibility, and originality of ideas (Carroll, 1993; French, Ekstrom, & Price, 1963). Hence, fluency and originality can be understood as aspects of DT. The relation of these two aspects with DT is discussed in the later sections. Although these aspects were often surmised, the literature still lacks psychometrically sound evidence for them. There have been several attempts of modelling DT as a general factor—mostly over fluency and originality—but these analyses have often fallen short of empirical validation (Carroll, 1993; Dumas & Dunbar, 2014; Kim, 2006). We will now define these essential aspects of DT and describe the current state of research regarding their measurement and relations with other constructs.

Fluency. Fluency captures the quantity of ideas and reflects the ability to produce a number of responses to a given problem (French et al., 1963). The ability to come up with a variety of answers has been classified as broad retrieval ability within the *three-stratum theory* of cognitive abilities (Carroll, 1993). It has been argued that retrieval ability and fluency are strongly contingent on cognitive or clerical speed (Forthmann, Jendryczko, et al., 2019) and that speediness biases DT scores (Forthmann, Szardenings, & Holling, 2018). Some researchers consider fluency an essential part of DT. From this perspective, a fluency/flexibility factor can be subsumed below broad retrieval ability (Silvia et al., 2013) and can explain over half of the variance in DT (Benedek et al., 2012). Because the quantity of ideas is easily measured and scored, most such DT tasks are reduced to a single fluency factor (Benedek, Jauk, Sommer, Arendasy, & Neubauer, 2014; Preckel, Wermer, & Spinath, 2011). Hence, a multitude of previously reported results are restricted to fluency scores only. However, simply equating fluency with DT is inadequate because it completely ignores the quality of an idea (e.g. Acar, Burnett, & Cabra, 2017).

Originality. Originality stresses the quality of ideas and evaluates how clever and uncommon they are

(Abraham, 2018). Originality indicates the cleverness, uncommonness, uniqueness, appropriateness, and usefulness of ideas on a prespecified topic (Carroll, 1993). Thus, originality resembles a key part of the consensus definition of broad creativity. This definition stresses the importance of originality in DT. Previous studies were inconclusive concerning the importance of originality. For one, the Educational Testing Service has decided to drop originality from its *Kit of Reference Tests for Cognitive Factors*, arguably because of its unclear status in the literature and unsuccessful efforts to develop suitable tasks for this factor (Ekstrom, French, Harman, & Derman, 1976). At this time, originality was just not well established in the research literature despite the work of Guilford. Nonetheless, many researchers see it as the most important ingredient of creativity (e.g. Acar et al., 2017). This view is strengthened by one of the most comprehensive factor analytic reviews of human abilities reporting a factor of originality (Carroll, 1993). Previous research indicates that fluency and originality are highly correlated on the manifest level ($r = .73$; Jung et al., 2015) and the latent level ($r = .89$; Silvia, 2008a). Because of high correlations between the scores, some researchers argue that the two are redundant and that originality can be dropped as it is easy and straightforward to measure fluency, but difficult and effortful to assess originality (e.g. Batey, Chamorro-Premuzic, & Furnham, 2009; Preckel, Holling, & Wiese, 2006). Other researchers conclude that originality is theoretically necessary and statistically distinct from fluency (Acar et al., 2017; Carroll, 1993; Dumas & Dunbar, 2014; Jauk, Benedek, & Neubauer, 2014), particularly when participants are carefully instructed (i.e. a 'be-creative' instruction; Nusbaum, Silvia, & Beaty, 2014). In summary, further investigation is encouraged by the conflicting theoretical considerations and the empirical evidence for a separable originality factor being a distinct dimension of DT. Additional robust evidence is required to understand whether originality can be established as a factor and whether such a factor adheres to expectations concerning convergent and discriminant validity.

How is divergent thinking scored?

Instruction and scoring of DT tasks are crucial, and variations in both most likely lead to diverging substantial results in terms of associations (Harrington, 1975; Nusbaum et al., 2014). The most common instruction and scoring of DT tasks stress verbal fluency. This means that only the quantity of responses is scored, resulting in a count variable. The literature provides a variety of scoring approaches to score the originality/quality of a response (Benedek, Muhlmann, Jauk, & Neubauer, 2013; Silvia et al., 2008). These scorings require much more complex human ratings, which are often associated with lower interrater reliability. Prior research suggests several ways to score the quality of an answer to obtain an originality score (scoring along the dimensions of uncommonness, remoteness, and cleverness: Cropley, 1967; Forthmann et al., 2017; Hocevar, 1979; Silvia, Martin, & Nusbaum, 2009; Vernon, 1971; Wilson, Guilford, & Christensen, 1953). On one hand, such traditional scoring methods based on the uniqueness of an answer may (i) be confounded with scores of fluency and originality, (ii) be biased in small sample sizes as the responses are not exhaustive, and (iii) yield many rare responses that are ambiguous in their interpretation (Silvia et al., 2008).

On the other hand, a simple aggregation over various DT tasks can be seen as problematic (see Reiter-Palmon, Forthmann, & Barbot, 2019). The literature provides various aggregation methods beyond sums, such as ratio quality scores (Forthmann et al., 2018) and residual scores (Runco & Albert, 1985). However, most aggregation scores suffer from low reliability, which can be attributed to confounding originality and fluency (Reiter-Palmon et al., 2019). Controversies about the reliability and validity of traditional scoring approaches (Benedek et al., 2013) resulted in the proposal of new scoring methods, such as the subjective top-2 method (Silvia et al., 2008). These approaches often yield lower correlations between fluency and originality, but also have downsides because participants are instructed to be fluent but then have to choose their two most original solutions, even though they were never instructed to be particularly original. Moreover, assessing the originality of answers is a challenging task, even for trained raters that have access to all answers given: this selection seems biased for participants who only have access to their own set of answers. Hence, an unequivocal instruction stressing originality and evaluating the single most-creative answer is arguably the best solution to avoid statistical dependencies and the confounding of fluency and originality in a single task (Nusbaum et al., 2014). An option for scoring such tasks provides the Consensual Assessment Technique (Amabile, 1982) that has often been described as the gold standard and can be used for any type of creativity ratings (Kaufman, Baer, Cropley, Reiter-Palmon, & Sinnett, 2013). Previous studies revealed that experts deliver highly reliable ratings using the Consensual Assessment Technique (Amabile, 1982; Kaufman et al., 2013).

How does creativity relate to intelligence, personality, and insight?

The relationship between creativity and other constructs (i.e. discriminant and convergent validity) is a key issue in research on creativity. DT has been linked to personality traits (i.e. openness; McCrae, 1987) and to intelligence (e.g. Kaufman & Plucker, 2011). Besides, insight has been linked to DT as well as to intelligence (Sternberg & O'Hara, 1999).

In the next paragraphs, we summarize knowns and unknowns in the relations of creativity with established intellectual abilities, insight, and personality traits to provide an integrative view of intelligence, personality, and creativity.

General intelligence. With respect to cognitive abilities, we stick to widely accepted models of human cognitive abilities (Carroll, 1993) and select key factors from this model. With the focus on DT, we understand creativity in terms of the general creative ability of a person, as described in the section 'What is Creativity?' A huge body of literature conceptualizes creativity based on indicators that stress maximal cognitive effort (Silvia et al., 2013)—just as any intelligence test does (Wilhelm & Schroeders, 2019). Although, creativity can be distinguished from intelligence: performance appraisal in the latter is based on a single and clearly correct answer, whereas creative performance is mostly assessed with open-ended answers that are rated by experts regarding their creativity. Historically, creativity and DT (or creative thinking; Guilford, 1956; Wilson et al., 1953) were embedded in various models of intelligence (e.g. *structure of intellect model*; Guilford, 1956), often considered as a lower level factor of it (e.g. active idea production in the three-stratum theory, including fluency and originality, subsumed under broad retrieval ability; Carroll, 1993). In sum, the intelligence literature provides models subsuming creativity as an ability factor below general intelligence (Carroll, 1993; Jäger et al., 1997; Süß & Beauducel, 2005). Intelligence and creativity are also closely intertwined within creativity research (Runco, 2004; Silvia, 2015; Silvia et al., 2013). For example, recent research (Silvia et al., 2013) provided support for the notion that originality and fluency, as components of DT, are subsumed by a broad retrieval factor (Schneider & McGrew, 2018). The relationship between intelligence and creativity, however, might not be so straightforward. On a meta-analytical level, the relationship between intelligence and creativity is rather low [$r = .17$; 95% CI (confidence interval) [.17, .18]; $N = 45\,880$; Kim, 2005].

Working memory and mental speed. Besides general intelligence, working memory capacity (a cognitive system needed to maintain and update mental representations; Oberauer, 2009) and mental speed (the speed of processing information; Danthiir, Wilhelm, Schulze, & Roberts, 2005) have also been studied in relation to creativity. Working memory updating ($\beta = .29$) and common executive functions like inhibition ($\beta = .20$) significantly predicted DT to a small extent (Benedek et al., 2014). Studies on the association between mental speed and DT have reported inconclusive results ranging from negative relations between reaction times and creativity (hick tasks: $r = -.18$; Vartanian, Martindale, & Kwiatkowski,

2007) to positive relations when tasks require inhibition (negative priming: $r = .28$, Vartanian et al., 2007) and to large positive correlations ($r = .63$; Vock, Preckel, & Holling, 2011). These inconclusive results indicate that the relation diminishes if models control for general intelligence. Alternatively, varying measures, scoring procedures, and instructions seem to play an important role. For example, it could be argued that only variance in fluency can be explained by mental speed, whereas individual differences in the quality of ideas might be independent of mental speed (Carroll, 1993; Forthmann, Jendryczko, et al., 2019).

Insight. Previous work highlighted the theoretical similarities between creativity and insight (Kounios & Beeman, 2014; Martindale, 1999; Schooler & Melcher, 1995). Insight (or eureka) moments are mostly based on the sudden recognition of a previously unknown conceptual connection followed by finding a new solution to a problem (e.g. Ball, Marsh, Litchfield, Cook, & Booth, 2015; DeCaro, 2018; Sprugnoli et al., 2017). Therefore, the similarity between creativity and insight is driven by reorganizing elements (e.g. words or pictures) and breaking existing patterns (Mednick, 1962, 1968). The literature also provides evidence for similarities between insight and intelligence. Insight tasks require maximal effort that relies on convergent thinking leading to a single and arguably veridical answer. Despite these conceptual similarities between creativity, insight, and intelligence, the empirical evidence shows small relations between insight (e.g. compound word associations) and creativity ($r = .28$, Mourgues, Preiss, & Grigorenko, 2014; $r = .31$, DeYoung, Flanders, & Peterson, 2008) and between insight and working memory capacity/intelligence ($r = .32/.44$, DeYoung et al., 2008).

Historically, Gestalt-psychological problems were used to study insight performance (e.g. Koehler, 1967) followed by approaches such as the nine-dot problem, the Duncker candle task, and the triangle of coins (for an overview, see Chu & MacGregor, 2011). All of these tasks provoke problem representations that do not allow for the application of well-practiced solutions. These tasks have major limitations, such as predominantly high item difficulties, poor time efficiency, low fidelity, and task heterogeneity with respect to stimuli and problems (Sprugnoli et al., 2017).

Because of its unclear status regarding the relations with other constructs and due to measurement problems, our study adds insight to the nomological net, as it shows conceptual overlap with intelligence and also creativity. In order to investigate these relations, we selected verbal insight problems (anagrams and riddles; Novick & Sherman, 2003) to reduce measurement limitations. Anagram and scrabble tasks rely on monitoring the constant process of assembling and

disassembling potential solutions and their matching with information retrieved from long-term memory.

Personality traits. Creative individuals seem to hold several relatively stable behavioural and personality characteristics that are associated with creative behaviour or result in creative products (Eysenck, 1993; Guilford, 1950; Sternberg & Lubart, 1993). Personality is often described by the five-factor model that suggests openness, conscientiousness, extraversion, agreeableness, and neuroticism as overarching traits (McCrae & Costa, 1989). Alternatively, the HEXACO model includes a sixth factor capturing honesty–humility (Ashton & Lee, 2001) and a different conceptualization of neuroticism and agreeableness compared with the five-factor model (Moshagen, Hilbig, & Zettler, 2014).

In both frameworks, several personality traits and facets are related to creativity (e.g. openness to novel experiences is associated with unconventional preferences, increased aesthetic sensibility, and attraction to complexity). Such relations between aspects of creativity and specific personality traits have been shown in numerous studies, although the magnitude of this relationship varies substantially. Previous approaches relating the five-factor model with creativity indicates that openness and its underlying facets, specifically fantasy, curiosity, and flexibility, are associated with several measures of creativity (Batey & Furnham, 2006; Feist, 1998) and are especially linked to DT (McCrae, 1987). On a trait level, openness has been demonstrated to be moderately associated with originality ($r = .26$) and fluency ($r = .31$; Jauk et al., 2014). A larger systematic review (Puryear, Kettler, & Rinn, 2017) supported the correlation between creativity and openness ($r = .24$; 95% CI [.23, .25]; $N = 57$ 019) and also found small correlations with extraversion ($r = .14$; 95% CI [.13, .15]; $N = 58$ 804), whereas the correlation with the other Big Five traits were close to zero, that is, for conscientiousness ($r = .02$; 95% CI [.01, .02]; $N = 58$ 897), agreeableness ($r = .03$; 95% CI [.02, .03]; $N = 57$ 068), and neuroticism ($r = -.04$; 95% CI [−.05, −.03]; $N = 56$ 748). Overall, openness and creativity seem to consistently have a small to medium relation (McCrae, 1987; Puryear et al., 2017), but other personality traits revealed a more diverse picture (e.g. extraversion), which is mainly due to the use of different assessment methods of creativity (e.g. self-report versus DT; Kandler et al., 2016; Puryear et al., 2017).

A study building upon the *HEXACO-60* did not find any significant relation between agreeableness and creativity ($\beta = -.04$) and conscientiousness and creativity ($\beta = -.04$; Silvia, Kaufman, Reiter-Palmon, & Wigert, 2011). It did however uncover a small negative association between honesty–humility and creativity ($\beta = -.20$), a small relation between creativity and extraversion ($\beta = .17$), and a strong association between openness and creativity ($\beta = .55$; Silvia et al., 2011).

Creativity, intelligence, and personality. Key theories of creativity stressed the importance of maximal cognitive effort and typical behaviour for creativity (e.g. Eysenck, 1993; Guilford, 1950). Despite numerous studies, the question to what extent creativity is distinct from other constructs of ability and personality is still unsolved. Arguably, the creative ability of a person interplays with individuals' personality and convergent thinking, as for example, DT in a specific domain requires knowledge in that area (Cropley, 2006). Various ways of measuring creativity (e.g. with objective measures, self-ratings, or other ratings; Batey, 2012) lead to diverging results. Importantly, measures of DT capture the part of maximal effort of creativity and are often more strongly related with other measures of maximal cognitive effort, whereas self-report measures of creativity are more akin to other measures of typical behaviour, arguably because of common method variance (Kandler et al., 2016).

Previous theories have outlined dependencies between personality and intelligence. The theory of adult intellectual development (Ackerman, 1996), for example, describes the incorporation and interaction of intelligence-as-process (fluid intelligence), personality, interests, and intelligence-as-knowledge. Ackerman assumes a substantial relation between openness and knowledge (crystallized intelligence), which has also been indicated in empirical studies (e.g. Ashton & Lee, 2001; von Stumm & Ackerman, 2013). To understand the empirical overlap between multiple constructs (such as creativity, intelligence, and personality), multivariate studies based on a variety of sound measures are necessary to draw a more complete picture of the nomological net (Ackerman, 2009) and to probe the uniqueness of creativity within it.

The present studies

We added two comprehensive multivariate studies to the existing body of research, including various creativity tasks as well as sound measures for cognitive abilities, personality traits, and insight. Although previous studies have also applied confirmatory factor analysis and embed creativity into a larger nomological net (e.g. Jauk et al., 2014), these studies are often only based on narrow measurement of DT. In the present paper, we address the dimensionality of DT including a large variety of tests measuring different aspects of DT, cognitive abilities, personality traits, and insight with a latent variable approach. In more detail, we model DT based on fluency and originality indicators, cognitive abilities (including fluid intelligence, crystallized intelligence, working memory, and mental speed), insight (anagrams and scrabbles), and personality traits (openness, extraversion, and

honesty–humility) in a confirmatory factor analytical framework. Our research objectives are (i) to assess the dimensionality of DT, including indicators of fluency and originality and (ii) to study the nomological net of DT by considering established cognitive abilities, personality traits, and insight with the above-mentioned factors. The research objectives and hypotheses were not preregistered.

In Study 1, we evaluated dimensions of DT by estimating and comparing a series of competing measurement models. DT was measured with two verbal, one figural, and one retrieval fluency tasks, as well as two originality tasks. The model series started by testing a general DT factor model (Model A), which was compared with a model estimating two correlated factors (originality and fluency; Model B) and a higher order factor model (Model C). The last model (Model D) was a nested factor model including a general DT factor and a nested originality factor. Model D tested the expectation that systematic individual differences reflecting originality exist above an overarching DT factor. In Study 2, we replicated the model series described earlier. The best-fitting measurement model of DT was used to study the nomological net of established cognitive abilities, personality traits, and insight. In line with the literature, we expected a moderate association with intelligence (including working memory and mental speed) and a small association with crystallized intelligence. Besides, insight is expected to be moderately related with DT as well as with general intelligence and crystallized intelligence. With respect to personality traits, we expected small positive associations with openness and extraversion, as well as a small negative association with honesty–humility. Based on the above-mentioned theoretical considerations, crystallized intelligence might mediate the relation between openness and DT.

Method

In the following section, we provide all information regarding the design and all measures that are applied in both studies. The sample size and all data exclusion criteria are described in detail in the following methods sections. For both studies, we did not determine the sample size in advance, but rather gathered participants in a given time slot until a meaningful sample size for confirmatory factor analysis was reached. Two other papers using a dataset that shows small overlaps with the dataset we used in Study 2 are submitted or accepted for publication.[2] All data needed to reproduce any of the reported results for both studies are available along with the syntax for statistical analysis at https://osf.io/c8j29/.

Procedure and design

The reported studies were conducted in three German cities (Greifswald, Ulm, and Bamberg). The test battery applied in Greifswald is described as Study 1. The test batteries used for Study 2 in Ulm and Bamberg were completely congruent and partly overlapping with the battery that was used in Greifswald. Because Study 2 aimed to validate and extend the measurement part established in Study 1, additional tasks were applied in Ulm and Bamberg.

Study 1. In Study 1, the test battery included a 2-hour behavioural assessment session with DT indicators and other covariates (see the section *Measures*). The tasks were administered in a computerized lab session and were programmed in PSYCHOPY (Peirce, 2007). Figural creativity and reasoning were paper-pencil based. Because of modelling issues on the item level, the flexibility task four-word sentences from the verbal creativity test (Schoppe, 1975) was excluded from subsequent analysis. In more detail, this task was scored for flexibility, but showed insufficient reliability. In addition, the Remote Associates Task (Landmann et al., 2014) was discarded from the analyses as we encountered severe scoring issues after data collection (e.g. more than one possible answer).

Study 2

In Study 2, the participants were subscribed to 7 hours of testing divided into a 5-hour lab and a 2-hour online session that they completed in advance on their home computers. All measures from Study 1 were also applied in Study 2. Additionally, Study 2 included measures of insight, a broader knowledge test, personality questionnaires (see the Measures section), self-reported creativity, a good taste, faking, and overclaiming. All newly developed tests were first assessed in a pilot study. The measures for the lab sessions were programmed in PSYCHOPY and Unipark; the online session was implemented on SoSciSurvey. In a lab session, groups of on average eight participants were tested at the same time. We restricted detailed descriptions of measures to the ones relevant for the research questions raised earlier. Tests that were not subject to this study include self-reported creativity (Diedrich et al., 2018), a newly developed multiple choice test of creativity (good-taste task), verb generation task to measure semantic distance as creativity phenotype in analogical reasoning (Green, Kraemer, Fugelsang, Gray, & Dunbar, 2012; Prabhakaran, Green, & Gray, 2014), typing speed, a corpus-based vocabulary test, overclaiming, and faking ability.

Sample

Study 1. Participants were recruited through university mailing lists and announcements in public places. Participants with major neurological or psychiatric disorders were excluded from the sample in both studies. All participants ($N = 159$) provided written informed consent and received monetary

reimbursement for their participation. Our final sample, after data preparation ($N = 152$; 54% female, for the cleaning procedure), ranged in age from 18 to 33 years ($M_{age} = 23.4$ years, $SD_{age} = 3.8$). Out of the final sample, 137 participants had a high school diploma with a mean grade of 2.1, ranging from 1 to 3.5 (higher grades indicate better school performance). One hundred and forty-two participants reported German as their mother tongue.

Study 2. Participants were recruited, informed, and incentivized in the same manner as in Study 1. A total of $N = 298$ (72% female; $M_{age} = 24.5$, $SD_{age} = 5.1$, age range from 18 to 49 years) was analysed after data cleaning (see data cleaning section). In the final sample, 278 participants had a high school diploma with a mean grade of 2.1 (ranging from 1 to 3.5). Two hundred and eighty of them reported German as their native language.

Measures

Creativity tasks applied in Studies 1 and 2. We applied two tasks each for fluency and originality to assess verbal creativity. Additionally, the following two insight measures (i.e. anagrams and scrabble words) were only applied in Study 2. Descriptive statistics and reliability estimates [intraclass correlation coefficient (ICCs)] for all single items are presented in Table S1 (Study 1) and Table S2 (Study 2). The ICCs were estimated as proposed by Shrout and Fleiss (1979). Based on our particular study design, we have chosen an ICC (ICC3k) that reflects the fact that a fixed set of raters rated all items.

Verbal fluency. In the similar attributes (SA) and inventing names (IN) fluency tasks, participants were instructed to produce as many appropriate answers as possible within a given time period (60 seconds). The SA task (e.g. 'Name as many things that you can that are "uneatable for humans"') was based on items out of the verbal creativity test (Schoppe, 1975). The test consisted of six timed items (60 seconds per item). The IN task (e.g. 'Invent names for the abbreviation: "T-E-F"') task was also adapted from Schoppe (1975). The test included 18 items, each with a 30 seconds time limit. The tests were open-ended and hence required human coding, as with all applied DT tasks. Three independent human coders thus applied a typical fluency coding (amount of correct answers). Further details regarding the scoring are given in the statistical analysis section. The inter-rater reliability was very high in both studies (SA ranging from .96 to 1.00; IN ranging from .93 to .99). We aggregated the scores provided by the three independent reviewers, resulting in a single mean score per item. After that, all items were aggregated to derive a task score.

Figural fluency. The tasks for assessing figural fluency (e.g. 'Draw as many objects as you can based on a circle and a rectangle') were adapted from the *Berliner Intelligenzstruktur-Test für Jugendliche: Begabungs- und Hochbegabungsdiagnostik* (Berlin Structure-of-Intelligence test for Youth: Diagnosis of Talents and Giftedness; Jäger, 2006). We employed four figural fluency/flexibility items that were assessed using paper-pen tests, as they required participants to draw figures. Figural tasks were coded by three independent human coders as well. The applied coding procedure followed the recommendations of the test manual and reached high interrater reliabilities (ICCs between .95 and .99). Because of the scope of the paper, we included a figural fluency test score across all four items.

Verbal originality. In both originality tasks, nicknames and combining objects (CO), participants were instructed to provide a single answer that was very unique and original. Three human coders once again rated the different answers. All human raters were semi-experts regarding creativity, and all went through a training procedure prior to rating. Similarly to the Consensual Assessment Technique (Amabile, 1982), we employed (semi-)experts to rate participants answers. All raters were trained in a 4-hour session in which the data, its structure, and the scoring guidelines along with a definition of creativity were explained. Every single answer was rated by each rater on a five-point scale based on scoring guidelines (Silvia et al., 2008, 2009). More precisely, an answer was rated as very creative if it was unique/rare/novel (uncommon), remote, and unexpected (clever) in the sample (Silvia et al., 2008). The raters were instructed to rate the creativity in relation to the answers given by other participants. Absent or inappropriate answers were coded as zero. Missing values in single tasks were due to computer problems and were deemed to be missing completely at random [Study 1: $n_{max} = 8$ (5.3%), $n_{mean} = 5.50$ (3.6%); Study 2: $n_{max} = 14$ (4.7%), $n_{mean} = 5.11$ (1.7%)]. During the rating procedure, the raters evaluated their responses independently and were only given the responses of the task they were currently rating. After collecting the ratings, we calculated ICCs and a compound score across all three raters for every item. The ICCs for originality were lower compared with the fluency scorings, but still acceptable (CO: ranging from .66 to .86; inventing names: ranging from .81 to .90), as expected. The items for CO (e.g. 'Combine two objects in order to build a door stopper in your house') were adapted and translated from English to German language from the Kit of Reference Tests for Cognitive Factors (Ekstrom et al., 1976). The task consisted of 12 items (with a time limit of 60 seconds per item). The nicknames items (e.g. 'Invent a combining objects for a bathtub')

were adapted from Schoppe (1975), including nine items with a 30-second time limit each.

Retrieval fluency. We adapted and translated six items (with a time limit of 60 seconds per item) for retrieval fluency tasks from the Kit of Reference Tests for Cognitive Factors (Ekstrom et al., 1976). Participants were asked to name as many things as they can in a given category. The categories were animals (e.g. dogs and birds) or household items. The answers were rated by two independent human raters. The agreement between the two raters was very high (ICC: .97 to 1.00). Hence, we aggregated the ratings into a sum-score based on the two ratings.

Cognitive ability tasks applied in Studies 1 and 2

Fluid intelligence. Fluid intelligence was assessed using figural (gf_f) and verbal (gf_v) reasoning tasks of the *Berlin Test of Fluid and Crystallized Intelligence* (Wilhelm, Schroeders, & Schipolowski, 2014). The verbal aspect of fluid intelligence was measured by tasks for relational reasoning. Its figural aspect required participants to infer how a sequence of geometric drawings—varying in shading and form according to certain rules—should continue. Each scale contained 16 multiple-choice items administered in order of increasing difficulty, with a 14-minute time limit per scale. For the verbal reasoning scale, the last two items were removed from the analysis because only a small proportion of participants solved them.

Working memory. As a working memory task, we applied a Recall-1-Back task including verbal (WM_v) and figural (WM_f) stimuli (Schmitz, Rotter, & Wilhelm, 2018; Wilhelm, Hildebrandt, & Oberauer, 2013). In the WM_v, task letters were displayed within a 3×3 matrix. Participants were instructed to type in the letter that appeared last in the matrix at a given position while remembering the current stimulus. Participants were thus asked to identify the position where the same symbol occurred last (see also Wilhelm et al., 2013). The task included a training phase with 21 trials and test phase including 66 classifications.

Mental speed. As mental speed tasks, we applied a computerized version of the comparison task (Schmitz & Wilhelm, 2016). In line with the creativity and the reasoning tasks, we applied verbal (MS_v) and figural (MS_f) stimuli. Participants were instructed to decide as quickly as possible whether two simultaneously presented triples of figures or letters on the screen were identical. The task consisted of two blocks of 40 trials each. As an indicator of mental speed, we used a reciprocal reaction time score (correct answers per time).

Further tasks applied only in Study 1

Crystallized intelligence. In Study 1, we assessed crystallized intelligence based on a 32-item short form of the *Berlin Test of Fluid and Crystallized Intelligence* (Wilhelm et al., 2014). We included knowledge items from three broad knowledge domains: natural sciences (gcnature), humanities (gchuman), and social studies (gcsocial). For the confirmatory factor analysis, items were parcelled according to their domain.

Further measures applied exclusively in Study 2

Insight tasks. Anagrams and scrabble tasks are measures of insight (Novick & Sherman, 2003; Schoppe, 1975; Sprugnoli et al., 2017). We developed one anagram and two scrabble tasks that were explicitly applied in an originality and fluency condition. The scrabble task with 14 items was applied with an originality condition (SCRorg; e.g. name the most original word that you can build out of a given word). Another scrabble task (including14 items) was applied in a fluency condition (SCRflu; e.g. provide as many words as you can think of). The two scores reflect independent tasks. There were 18 items in the anagram task (ANAorg), all applied with an originality instruction (e.g. name the most creative anagram you can think of). Both conditions have a small number of correct solutions (e.g. three correct solutions for a given anagram) and require a certain degree of crystallized intelligence and general intelligence. Hence, we think that the ability to even produce a creative anagram out of this smaller response spectrum diverges from the ability needed in other originality and fluency conditions. Therefore, we refrained from subsuming these tasks below factors designed to capture communality of traditional originality and fluency measures and modelled insight as a unique factor. These three tasks were only administered in Study 2 and were also scored by three independent human raters. The fluency conditions were scored for the quantity of correct responses, and the originality conditions were scored—in line with the Consensual Assessment Technique—for the quality of a response. Table S3 shows the descriptive statistics (means and standard deviations) along with the reliability (ICCs) for all items.

Crystallized intelligence. Crystallized intelligence was assessed as declarative knowledge by two parallel test forms of a knowledge quiz with 136 items each, covering questions from natural sciences (gcnature), life sciences (gclife), social sciences (gcsocial), humanities (gchuman), and pop culture (gcpop). Participants randomly received either version A or B of the test in the online assessment and subsequently the other version in the lab session. Questions were sampled from a larger item pool of multiple-choice items (Steger, Schroeders, & Wilhelm, 2019) and selected according to content and difficulty. Here, we only analysed the

items applied in the proctored lab session as an unbiased indicator for crystallized intelligence. In the confirmatory factor analysis, we included parcels for the four broad knowledge domains of natural sciences, humanities, social studies, and life sciences. The pop culture items were dropped from the analysis as they covered current events knowledge, which differs from more traditional academic knowledge taught in schools (Beier & Ackerman, 2001).

Personality traits. To assess narrow-sense personality traits, we used the German 60-item version of the HEXACO (Moshagen et al., 2014), covering the personality traits of honesty–humility, emotional stability, extraversion, agreeableness, conscientiousness, and openness. Because of the previous results, only honesty–humility, extraversion, and openness were related with DT. For the measurement models, we used three parcels per personality trait. Because of unacceptable fit of the measurement model, we decided against using the HEXACO-60 facets as parcels (Ashton & Lee, 2009). Instead, we used three homogenous parcels with similar mean values that included the facets randomly. In Tables 1 and 2, we report descriptive statistics, reliability estimates, and correlations (including exact *p* values) for the measures that were analysed in both studies. Table 2 also includes the relations for all indicators with the six personality traits measured by the HEXACO-60. In line with previous research (e.g. Silvia et al., 2011), emotionality, agreeableness, and conscientiousness were not significantly associated with any of the creativity indicators, except a small correlation between conscientiousness and figural fluency and retrieval fluency that becomes nonsignificant if adjusted for multiple testing (based on the Holm's method; Holm, 1979).

Data preparation

Participants were excluded if they were older than 50 years, as an older age is clearly associated with age-related decline and larger variability in cognitive functions across persons (Hartshorne & Germine, 2015). Ninety-five per cent of both samples consisted of participants with higher educational degrees. To homogenize the sample and to remove multivariate outliers, we decided to exclude all participants ($n = 24$) with lower educational degrees (of vocational-track *Hauptschule* schools or no school degree). During data cleaning, we excluded participants deemed multivariate outliers across all DT tasks based on the Mahalanobis distance (see also Meade & Craig, 2012). The Mahalanobis distance is the standardized distance of one data point from the mean of the multivariate distribution. Following these steps of data cleaning, we excluded seven participants from the sample collected in Study 1 and 17 participants from the sample of Study 2.

Measurement models

To reduce model complexity, we decided to use test scores of DT as indicators in the confirmatory factor models. To justify the usage of a test score, we first tested for unidimensionality by fitting measurement models on item level for all DT tests described earlier. The measurement models on the task-level are provided in the Supporting information (Study 1: Table S4 and Study 2: Table S5). All unidimensional measurement models reached acceptable to very good fit except the measurement model of retrieval fluency in Study 2. Therefore, we used manifest test scores as indicators in all subsequent analyses. As described earlier, we computed a mean across all raters for every item. The test scores were then based on the mean value across all items. The correlations between the manifest variables based on sum scores of DT (fluency and originality) with fluid and crystallized intelligence (Studies 1 and 2), insight, openness, extraversion, and honesty–humility (only Study 2) are displayed in a scatterplot in Figures S1A and S1B. For evaluating model fit, we used the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) (Hu & Bentler, 1999). Conventionally, CFI $\geq .95$, RMSEA $\leq .06$, and SRMR $\leq .08$ indicate very good fit (Hu & Bentler, 1999). Analyses were conducted with the R software (version 3.6.2) using the packages *lavaan* (Rosseel, 2012) for latent variable modelling and *psych* (Revelle, 2018) for descriptive statistics. All confirmatory models were estimated with the *maximum likelihood* (ML) estimator and *full information ML* estimator to handle missing values. Full information ML is a state-of-the-art method in structural equation modelling (Schafer & Graham, 2002) because it retains the statistical power and allows for more precise parameter estimation in comparison with traditional missing data treatment methods (Enders, 2010). As a reliability estimate, we used McDonald's ω (McDonald, 1999) because it assumes a tau-congeneric measurement model (Raykov & Marcoulides, 2011). The factor saturation (ω) for a factor indicates how much variance is accounted for by a latent variable in all underlying indicators (Brunner, Nagy, & Wilhelm, 2012).

Results

Study 1

In Study 1, we compared competing measurement models to address the dimensionality of DT. More precisely, we tested a series of models as schematically outlined in Figure 1. Model A assumes a general factor reflecting DT. Model B postulates two correlated factors (fluency* and originality*), Model C shows a higher order model (including two

Table 1. Means, standard deviations, and correlations including exact *p* values in parentheses for measures of Study 1.

	Mean (SD)	ω	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 Age	23.37 (3.78)		−.02 (.83)	.00 (.98)	**−.19 (.04)**	−.08 (.41)	.17 (.07)	.13 (.17)	−.15 (.11)	−.04 (.66)	**−.21 (.02)**	−.17 (.07)	−.05 (.57)	**−.20 (.03)**	**.27 (.00)**	.09 (.35)	**.31 (.00)**
2 Sex	1.46 (.50)		**−.24 (.01)**	**−.19 (.04)**	**−.28 (.00)**	**−.22 (.02)**	.11 (.22)	−.17 (.07)	−.18 (.06)	−.09 (.35)	**−.26 (.01)**	**−.23 (.01)**	−.02 (.79)	−.13 (.16)	**.23 (.01)**	.11 (.26)	**.25 (.01)**
3 SA	6.79 (2.15)	.84		**.39 (.00)**	**.54 (.00)**	**.48 (.00)**	**.39 (.00)**	**.27 (.00)**	.06 (.53)	.00 (.98)	.04 (.66)	−.02 (.87)	**.20 (.03)**	.18 (.06)	**.23 (.01)**	**.22 (.02)**	.14 (.13)
4 IN	.77 (.38)	.90			**.29 (.00)**	**.28 (.00)**	**.24 (.01)**	**.30 (.00)**	.06 (.54)	−.04 (.71)	.08 (.38)	.07 (.45)	.13 (.15)	.15 (.12)	**.19 (.04)**	**.23 (.01)**	.12 (.18)
5 FF	5.24 (1.43)	.70				**.35 (.00)**	.05 (.56)	**.19 (.04)**	.17 (.07)	.05 (.58)	**.24 (.01)**	.13 (.15)	**.27 (.00)**	**.22 (.02)**	.02 (.85)	.00 (.98)	−.06 (.51)
6 RF	8.29 (2.45)	.84					**.30 (.00)**	**.26 (.00)**	**.20 (.03)**	**.20 (.03)**	.15 (.12)	.17 (.07)	.11 (.25)	**.35 (.00)**	.15 (12.)	**.21 (.03)**	.05 (.56)
7 CO	1.80 (.40)	.64						.18 (.06)	−.10 (.30)	−.11 (.25)	−.08 (.42)	−.11 (.22)	.08 (.41)	.17 (.06)	**.38 (.00)**	**.27 (.00)**	.18 (.05)
8 NI	2.04 (.66)	.75							.10 (.27)	.07 (.44)	.00 (.99)	.09 (.33)	.16 (.08)	**.21 (.03)**	.12 (.21)	**.26 (.01)**	**.20 (.03)**
9 WM$_f$	.74 (.14)									**.27 (.00)**	.18 (.06)	.19 (.04)	**.31 (.00)**	**.38 (.00)**	−.06 (.54)	−.03 (.73)	−.02 (.83)
10 WM$_v$	.77 (.14)										.13 (.15)	.13 (.15)	**.27 (.00)**	**.33 (.00)**	−.02 (.83)	.03 (.73)	.06 (.49)
11 MS$_f$	.91 (.16)											**.79 (.00)**	.11 (.24)	.16 (.08)	−.15 (.11)	−.08 (38)	−.17 (.07)
12 MS$_v$	1.08 (.19)												.03 (.75)	**.22 (.02)**	−.09 (.33)	−.07 (.48)	−.08 (.38)
13 gf$_f$	.54 (.17)	.67												**.39 (.00)**	**.23 (.01)**	.13 (.17)	**.20 (.03)**
14 gf$_v$	.68 (.16)	.63													**.26 (.01)**	.11 (.25)	.14 (.12)
15 gc$_{social}$	.68 (.18)															**.39 (.00)**	**.41 (.00)**
16 gc$_{human}$	.81 (.15)																**.32 (.00)**
17 gc$_{nature}$	.78 (.14)																

Note: Statistically significant correlations (*p* < .05) are bold. Fluency indicators are SA (similar attributes), IN (inventing names), FF (figural fluency), and RF (retrieval fluency). CO (combining objects) and NI (nicknames) are originality indicators. gc, crystallized intelligence (social, social sciences; human, humanities; nature, natural sciences); gf, fluid intelligence (f, figural; v, verbal); MS, mental speed (f, figural; v, verbal); WM, working memory capacity (f, figural; v, verbal). ω as a reliability estimator is presented for unidimensional measurement models of each scale. ω_{WM} = .47, ω_{MS} = .83, and ω_{gc} = .65.

first-order factors, fluency[+] and originality[+], and one second-order factor of DT[+]), whereas Model D is set up to estimate a factor of originality[#] that is nested below a general DT[#] factor. It should be noted that factors with the same label have to be interpreted differently (marked with *, [+], and [#]). For example, DT (Model A) and DT[#] (Model D) vary with respect to the breadth of their measurement. Models B and C are equivalent as long as no covariates are added to the models (MacCallum, Wegener, Uchino, & Fabrigar, 1993), whereas Model C (higher order model) and Model D (nested factor model) are quite similar representations of the data (Reise, Moore, & Haviland, 2010). In general, a higher order model can be converted to a constrained version of a (complete) bifactor model with nested factors for the previously first-order factors (Mulaik & Quartetti, 1997) by means of the Schmid-Leiman decomposition (Schmid & Leiman, 1957). The

nested factor model we estimate (Model D) has been described as bifactor-(S-1) model in the psychometric literature (Eid, Geiser, Koch, & Heene, 2017; Eid, Krumm, Koch, & Schulze, 2018), because it contains only one specific factor (S) and omits the second factor, fluency, as a reference. This modelling approach avoids the usual problems of a bifactor model (or the higher order model, as a special version of the bifactor model) such as vanishing factors, negative variances, and irregular loading patterns (see Eid et al., 2017). Besides, the proportionality constraints that are applied by diverging a higher order model from the corresponding hierarchical model (Mulaik & Quartetti, 1997) are mostly of small theoretical value. Moreover, embedding the higher order model in the nomological net does not allow for a simultaneous estimation of all correlations between second-order and first-order factors and covariates because these relationships are linearly dependent

Table 2. Means, standard deviations, and correlations including exact p values in parentheses for measures of Study 2.

		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	age	−.03 (.68)	−.08 (.22)	**−.16 (.01)**	−.11 (.07)	−.11 (.08)	.03 (.66)	.02 (.70)	−.04 (.54)	−.09 (.14)	**−.15 (.02)**	**−.23 (.00)**	**−.39 (.00)**	**−.29 (.00)**	**−.28 (.00)**	**−.34 (.00)**
2	sex	**−.13 (.04)**	−.07 (.25)	**−.15 (.02)**	**−.21 (.00)**	−.02 (.87)	.09 (.18)	.00 (.84)	−.10 (.11)	.00 (.77)	−.02 (.57)	−.02 (.62)	**−.15 (.02)**	−.02 (.55)	**.13 (.04)**	−.02 (.77)
3	SA		**.45 (.00)**	**.47 (.00)**	**.48 (.00)**	**.47 (.00)**	**.26 (.00)**	.11 (.10)	**.31 (.00)**	**.21 (.00)**	.06 (.37)	**.25 (.00)**	.12 (.06)	.10 (.12)	**.23 (.00)**	.17 (.17)
4	IN			**.31 (.00)**	**.31 (.00)**	**.39 (.00)**	**.21 (.00)**	.07 (.25)	**.24 (.00)**	**.26 (.00)**	.10 (.12)	**.18 (.00)**	.04 (.53)	.05 (.40)	**.19 (.00)**	**.17 (.01)**
5	FF				**.37 (.00)**	**.35 (.00)**	**.16 (.02)**	.08 (.17)	**.22 (.00)**	.13 (.05)	.11 (.09)	**.23 (.00)**	**.22 (.00)**	.10 (.10)	**.25 (.00)**	**.17 (.01)**
6	RF					**.33 (.00)**	**.16 (.01)**	**.17 (.01)**	**.39 (.00)**	**.26 (.00)**	.07 (.33)	**.20 (.00)**	.12 (.05)	**.15 (.03)**	**.17 (.01)**	**.15 (.02)**
7	CO						**.33 (.00)**	.05 (.35)	**.15 (.01)**	**.17 (.00)**	.02 (.62)	**.18 (.00)**	**.15 (.02)**	.09 (.14)	**.23 (.00)**	**.20 (.00)**
8	NI							**.13 (.03)**	.11 (.09)	**.26 (.00)**	.01 (.93)	.06 (.35)	.02 (.74)	.09 (.16)	.08 (.22)	.06 (.27)
9	SCRorg								**.52 (.00)**	**.33 (.00)**	.12 (.07)	**.15 (.02)**	−.04 (.55)	.00 (.98)	**.19 (.00)**	.03 (.66)
10	SCRflu									**.46 (.00)**	**.20 (.00)**	**.23 (.00)**	.04 (.59)	.04 (.47)	**.25 (.00)**	.06 (.34)
11	ANAorg										**.19 (.00)**	**.29 (.00)**	.09 (.15)	.10 (.14)	**.27 (.00)**	.12 (.06)
12	WMf											**.35 (.00)**	**.13 (.04)**	−.04 (.54)	**.30 (.00)**	**.20 (.00)**
13	WMv												**.22 (.00)**	**.19 (.00)**	**.48 (.00)**	**.23 (.00)**
14	MSf													**.71 (.00)**	**.22 (.00)**	**.37 (.00)**
15	MSv														**.17 (.00)**	**.20 (.00)**
16	gff															**.43 (.00)**
17	gfv															

		Mean (SD)	ω	18	19	20	21	22	23	24	25	26	27
1	age	24.45 (5.10)		**.13 (.03)**	**.17 (.01)**	.10 (.11)	.10 (.10)	.11 (.07)	−.03 (.60)	−.12 (.06)	.02 (.71)	**−.14 (.03)**	−.01 (.89)
2	sex	1.28 (.46)		**.32 (.00)**	−.05 (.27)	.06 (.44)	**.20 (.00)**	−.08 (.21)	.08 (.24)	−.00 (.93)	.00 (.90)	**−.43 (.00)**	−.12 (.06)
3	SA	7.40 (2.10)	.79	**.18 (.01)**	**.28 (.00)**	**.22 (.00)**	**.19 (.00)**	.08 (.21)	.08 (.22)	.08 (.20)	−.10 (.12)	−.09 (.17)	.04 (.49)

(continued)

Table 2. Continued

		Mean (SD)	ω	18	19	20	21	22	23	24	25	26	27	28
4	IN	.82 (.37)	.91	**.22 (.00)**	**.20 (.00)**	**.27 (.00)**	**.22 (.00)**	.13 (.05)	.00 (.99)	.06 (.32)	.02 (.79)	.04 (.57)	−.10 (.11)	−.01 (.87)
5	FF	5.84 (1.66)	.65	**.19 (.00)**	**.16 (.01)**	**.15 (.02)**	.06 (.41)	−.02 (.81)	.12 (.07)	**.14 (.03)**	−.04 (.56)	−.04 (.57)	.05 (.39)	**.16 (.01)**
6	RF	8.42 (2.30)	.74	.11 (.11)	**.33 (.00)**	**.32 (.00)**	**.20 (.00)**	**.15 (.02)**	.03 (.65)	**.14 (.02)**	−.06 (.30)	.07 (.23)	.01 (.85)	**.15 (.02)**
7	CO	1.77 (.38)	.61	**.20 (.00)**	**.25 (.00)**	**.22 (.00)**	**.18 (.01)**	**.13 (.04)**	.12 (.07)	.09 (.13)	−.07 (.29)	−.03 (.65)	.04 (.52)	.05 (.46)
8	NI	1.91 (.69)	.81	**.20 (.00)**	**.25 (.00)**	**.27 (.00)**	**.18 (.00)**	**.16 (.01)**	**.18 (.01)**	**.13 (.04)**	−.04 (.56)	.03 (.64)	−.07 (.27)	.01 (.85)
9	SCRorg	1.76 (.60)	.81	.10 (.12)	.07 (.28)	**.16 (.01)**	−.01 (.94)	.07 (.30)	.08 (.19)	.00 (.93)	.04 (.50)	.06 (.36)	.05 (.45)	−.01 (.90)
10	SCRflu	2.54 (.79)	.89	**.20 (.00)**	**.16 (.01)**	**.14 (.03)**	.10 (.12)	.09 (.18)	−.02 (.66)	.02 (.80)	−.01 (.84)	.06 (.32)	.02 (.78)	.01 (.89)
11	ANAorg	.59 (.29)	.76	**.27 (.00)**	**.25 (.00)**	**.27 (.00)**	**.17 (.01)**	.12 (.05)	.11 (.09)	.00 (.96)	.12 (.06)	.11 (.10)	−.04 (.57)	.05 (.43)
12	WM$_f$	.74 (.12)		.12 (.05)	.01 (.92)	.09 (.17)	.05 (.37)	.02 (.72)	−.02 (.65)	.05 (.39)	−.05 (.41)	.05 (.41)	.00 (.98)	.11 (.07)
13	WM$_v$	.77 (.17)		**.31 (.00)**	.12 (.07)	**.13 (.04)**	.08 (.20)	.08 (.20)	−.01 (.85)	.03 (.62)	−.04 (.53)	.12 (.06)	.05 (.47)	**.18 (.00)**
14	MS$_f$	.89 (.16)		**.14 (.03)**	.01 (.92)	−.05 (.41)	**−.14 (.02)**	**−.14 (.02)**	.05 (.42)	.12 (.06)	−.06 (.32)	.06 (.38)	.12 (.06)	.09 (.15)
15	MS$_v$	1.05 (.18)		.09 (.17)	.01 (.90)	**−.14 (.03)**	**−.13 (.04)**	−.10 (.14)	.00 (.92)	.12 (.05)	−.04 (.53)	.06 (.34)	.12 (.05)	.04 (.57)
16	gf$_f$	.53 (.19)	.71	**.42 (.00)**	**.19 (.00)**	**.15 (.01)**	**.22 (.00)**	.07 (.31)	.02 (.71)	**.17 (.01)**	.00 (.95)	−.01 (.82)	.09 (.15)	**.14 (.03)**
17	gf$_v$	.69 (.18)	.69	**.22 (.00)**	.10 (.12)	.11 (.11)	**.21 (.00)**	.10 (.12)	−.07 (.24)	.12 (.05)	.02 (.73)	.00 (.98)	.00 (.90)	.06 (.30)
18	gc$_{nature}$	.58 (.15)			**.33 (.00)**	**.32 (.00)**	**.38 (.00)**	**.22 (.00)**	**.20 (.00)**	−.03 (.60)	.06 (.38)	−.08 (.21)	.11 (.07)	.05 (.42)
19	gc$_{life}$	.60 (.13)				**.33 (.00)**	**.38 (.00)**	**.35 (.00)**	.11 (.08)	.02 (.76)	.03 (.66)	−.05 (.40)	−.01 (.92)	.04 (.54)
20	gc$_{human}$	.58 (.13)					**.51 (.00)**	**.42 (.00)**	**.21 (.00)**	.03 (.66)	.03 (.60)	.00 (.99)	−.08 (.20)	.00 (.97)
21	gc$_{social}$	.54 (.17)						**.40 (.00)**	**.17 (.01)**	.03 (.69)	.04 (.53)	−.07 (.26)	−.09 (.16)	−.02 (.75)
22	gc$_{pop}$	.55 (.15)							.08 (.22)	−.02 (.75)	−.11 (.08)	**.17 (.01)**	**−.24 (.00)**	.07 (.27)
23	openness	3.45 (.69)	.75							**.22 (.00)**	.03 (.60)	.06 (.33)	**.15 (.02)**	.01 (.85)
24	extraversion	3.39 (.72)	.83								.00 (.99)	.06 (.39)	**.13 (.04)**	**−.15 (.02)**
25	honesty–humility	3.45 (.58)	.79									.03 (.68)	**.26 (.00)**	.10 (.11)
26	emotionality	3.25 (.77)	.85										−.07 (.25)	**.22 (.00)**
27	agreeableness	3.14 (.82)	.81											−.06 (.31)
28	conscientiousness	3.42 (.72)	.83											

Note: Statistically significant correlations ($p < .05$) are displayed in bold. Fluency indicators are SA (similar attributes), IN (inventing names), FF (figural fluency), and RF (retrieval fluency). CO (combining objects) and NI (nicknames) are originality indicators. gc, crystallized intelligence (social, social sciences; human, humanities; nature, natural sciences; life, life sciences; pop, pop culture); gf, fluid intelligence (f, figural; v, verbal); WM, working memory capacity (f, figural; v, verbal); MS, mental speed (f, figural; v, verbal). Scrabble words (SCR) and anagrams (ANA) are indicators of insight scored for originality (org) and/or fluency (flu). ω as a reliability estimator is presented for unidimensional measurement models of each scale. $\omega_{WM} = .52$, $\omega_{MS} = .82$, and $\omega_{gc} = .72$.
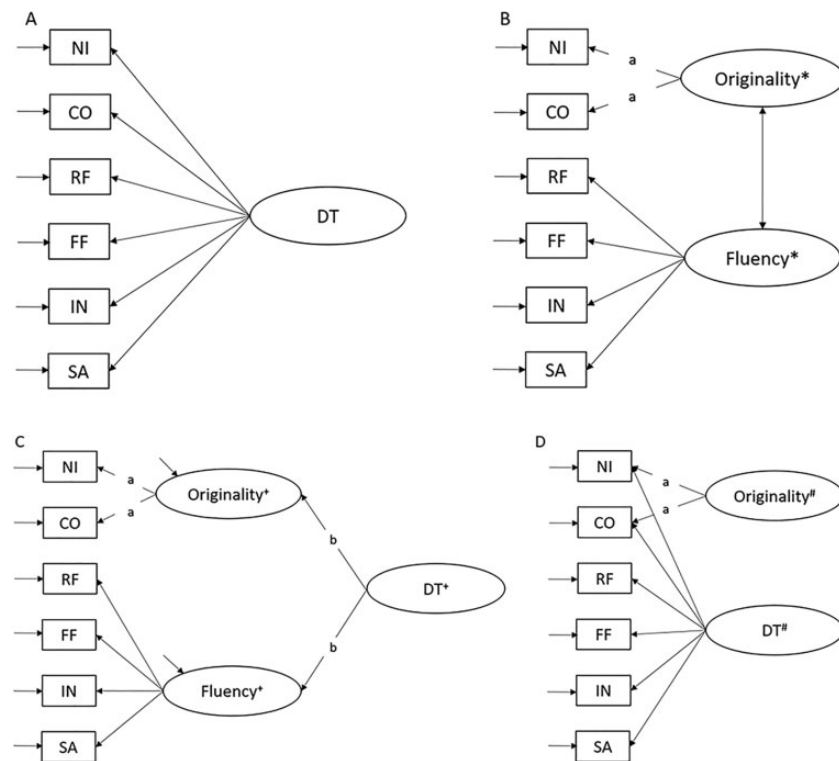
Figure 1. (A–D) Competing measurement models of divergent thinking (DT). Indicators are based on test scores computed as described in the method section. Fluency indicators are SA (similar attributes), IN (inventing names), FF (figural fluency), and RF (retrieval fluency). CO (combining objects) and NI (nicknames) are originality indicators that were only instructed for originality. *, +, and # indicate a different interpretation of the according latent factor compared with the other models. The factor variances of the latent variables were fixed to 1. All factors were scaled using unit variance identification constraints (Kline, 2015).

Table 3. Fit indices of the models displayed in Figure 1 as estimated in Study 1.

Model	$\chi^2(df)$	p	CFI	RMSEA [90% CI]	SRMR	ω
A	13.40 (9)	.15	.98	.06 [.00, .12]	.04	DT =.77
B	13.52 (9)	.14	.97	.06 [.00, .12]	.05	Fluency* = .75
						Originality* =.33
C	13.52 (9)	.14	.97	.06 [.00, .12]	.05	Fluency$^+$ =.75
						Originality$^+$ =.33
D	12.52 (8)	.13	.97	.06 [.00, .12]	.04	DT$^#$ = .77
						Originality$^#$ =.14

Note: ω, factor saturation (McDonald, 1999); CFI, comparative fit index; CI, confidence interval; RMSEA, root mean square error of approximation; SRMR, standardized root mean-square residual. *, +, and # indicate a change in factor composition.

(Schmiedek & Li, 2004). In sum, Model D prevents issues of collinearity, is less constrained, and allows for testing incremental contributions of covariates, which is why we prefer Model D over Model C. But keep in mind that the differences between both models are minor, and pursuing either a nested factor or higher order approach does not affect the conclusions we draw.

All models (A to D) had acceptable to good fit (see Table 3). Standardized loadings for all four models are presented in Figure S2. Because Models A and D are nested models, their fit was compared with a χ^2-difference test, which indicated that both models were not significantly different

($\Delta\chi^2(1, N = 152) = .89, p = .35$). In Model B, the correlation of the two latent factors (originality* and fluency*) was very high ($r = .82; p < .001$), as expected. The originality (originality$^#$) factor in the nested model (Model D) had very low factor saturation (as indicated by the ω coefficient in Table 3; McDonald, 1999), and its variance was inferentially not larger than zero ($p = .35$). These results indicated that all models fit the data similarly well. If the limited variance in the originality$^#$ factor is true, a larger sample will allow assessing its dispersion and saturation more precisely. Whether the originality$^#$ factor is a useful psychological construct could not be settled definitely in Study 1. Thus, we further examined this
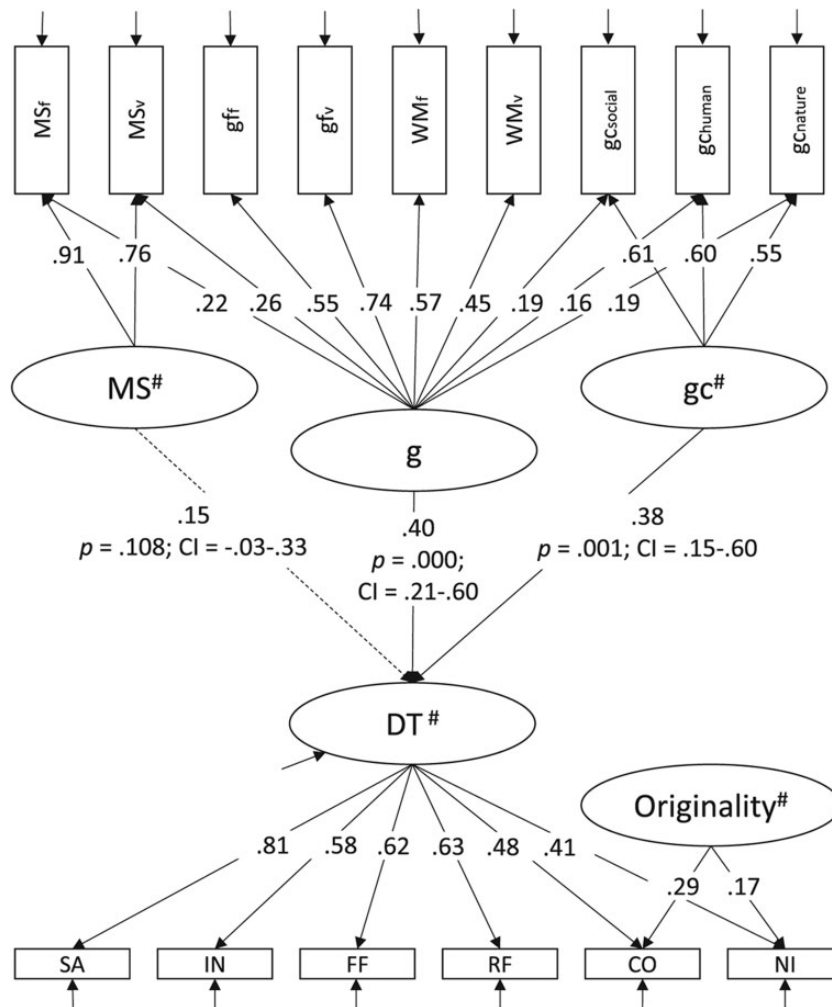
Figure 2. Structural model (Study 1; $N = 152$) relating DT$^\#$ to general cognitive ability (g), crystallized intelligence (gc$^\#$), and mental speed (MS$^\#$). Nonsignificant latent regressions are displayed as dotted lines. All coefficients are standardized. Nested factors in the cognitive ability model are MS$^\#$, specific factor of mental speed; gc$^\#$, specific factor of crystallized intelligence. The indicators of intelligence include test scores for figural (MS$_f$) and verbal mental speed (MS$_v$), fluid intelligence (figural, gf$_f$; verbal, gf$_v$); WM, working memory (figural, WM$_f$; verbal, WM$_v$), and parcels for gc in natural sciences, humanities, and social studies. The factor variances of the latent variables were fixed to 1. All factors were scaled using unit variance identification constraints (Kline, 2015). CI, confidence interval; CO, combining objects; FF, figural fluency; IN, inventing names; NI, nicknames; RF, retrieval fluency; SA, similar attributes.

theory-driven model (Model D) in Study 2 with a larger sample.

Before readdressing the analyses on the dimensionality of DT based on the larger sample available in Study 2, we embedded Model D into the nomological net of further cognitive abilities. Figure 2 illustrates the nested factor model of DT (Model D) together with cognitive abilities. This cognitive ability part of the model was based upon indicators of mental speed, working memory capacity, fluid intelligence, and three indicators of crystallized intelligence. The cognitive ability model assumed an overarching general factor of intelligence and a nested mental speed (MS$^\#$) and crystallized intelligence (gc$^\#$) factor. The fit of the model was good given the model complexity, although not optimal: $\chi^2(82) = 120.00$, CFI = .92, RMSEA = .06, SRMR = .07. DT$^\#$ was predicted by the g-factor and the orthogonal crystallized intelligence (gc$^\#$) factor: both had a moderate effect size

and explained 32% of the variance of the DT$^\#$ factor. A model including regressions between the originality$^\#$ factor and cognitive abilities in Study 1 led to estimation problems, most likely due to the limited sample size.

Study 2

In Study 2, we first reassessed the measurement models of DT based on the larger sample. Figure 1 displays the model series as estimated in both studies. Table 4 summarizes fit indices for the model series in Study 2. Models A and D were nested models; hence, the fit indices can be compared inferentially with a χ^2-difference test, which indicates that Model D was significantly better fitting than Model A [$\Delta\chi^2(1, N = 298) = 6.24$, $p = .01$]. In Model B, the correlation between originality* and DT* was high, as expected ($r = .79$; $p < .001$). The originality$^\#$ factor in the nested

Table 4. Fit indices of the models displayed in Figure 1 as estimated in Study 2

Model	$\chi^2(df)$	p	CFI	RMSEA [90% CI]	SRMR	ω
A	11.55 (9)	.24	.99	.03 [.00, .08]	.03	DT = .74
B	5.33 (9)	.81	1.00	.00 [.00, .04]	.02	Fluency* = .72
						Originality* = .42
C	5.33 (9)	.81	1.00	.00 [.00, .05]	.02	Fluency+ = .72
						Originality+ = .42
						DT+ = .62
D	5.31 (8)	.72	1.00	.00 [.00, .05]	.02	DT# = .75
						Originality# = .22

Note: ω, factor saturation (McDonald, 1999); CFI, comparative fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean-square residual. *, +, and # indicate a change in factor composition.
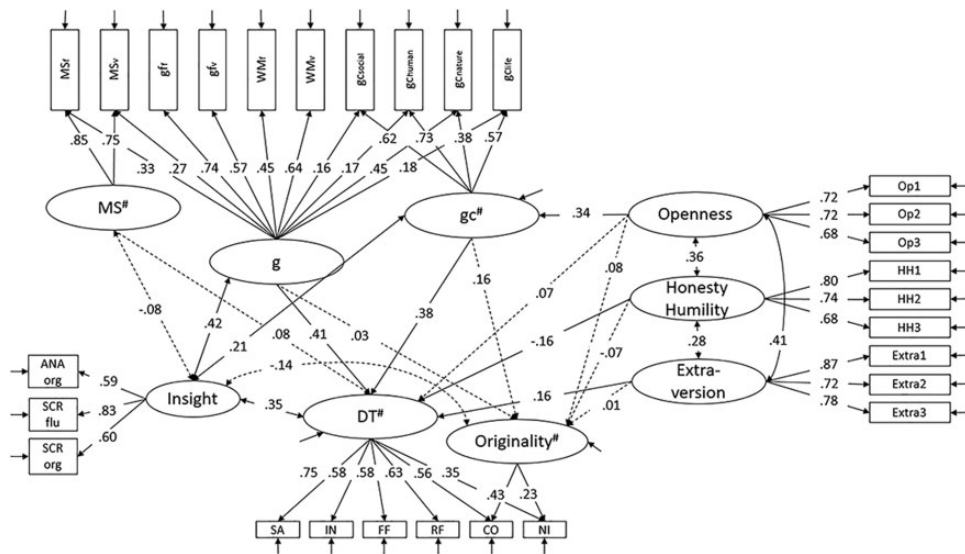
Figure 3. Structural model (Study 2; $N = 298$) relating DT# and originality# to general cognitive ability (g), gc#, and MS# insight and personality traits. Nonsignificant latent regressions are displayed as dotted lines. All coefficients are standardized. Nested factors in the cognitive ability model are MS#, specific factor of mental speed; gc#, specific factor of crystallized intelligence. The indicators of intelligence include test scores for figural and verbal mental speed (MSfig, MSverb), fluid intelligence (gffig, gfverb); WM, working memory (WMfig, WMverb); and parcels for gc in natural sciences, humanities, and social studies. Personality indicators are based on three parcels for the respective factor. ANA, anagrams; CO, combining objects; FF, figural fluency; IN, inventing names; NI, nicknames; RF, retrieval fluency; SA, similar attributes; SCR, scrabble words.

model (Model D) still possessed a low factor saturation, but nonetheless captured substantial variance ($p = .02$). Overall, the results suggested that Models B, C, and D fit the data similarly well.

Finally, we embedded the measurement model of Model D into the nomological net of cognitive abilities, personality traits, and insight. Figure 3 displays a model including all theoretically proposed relations. Additionally, we provide a table in Table S6 that provides all potential relations of the model displayed in Figure 3. However, allowing all relations did not substantially improve model fit. Moreover, the magnitude of the relations did not change [expect for small significant relations between extraversion and g and crystallized intelligence (gc#), respectively]. The fit of the model displayed in Figure 3 was good given the high model complexity, although not

optimal: $\chi^2(324) = 502.108$, $p < .001$, CFI = .93, RMSEA = .04, SRMR = .06. Exact p values along with 95% CIs of the relations displayed in Figure 3 are presented in Table S7. In sum, the results were comparable with the results of Study 1. General DT# was significantly predicted by the g-factor and the orthogonal crystallized intelligence (gc#) factor. The nested factor of mental speed (MS#) did neither predict DT#, nor originality#. As expected, insight was correlated with the g-factor and the orthogonal crystallized intelligence (gc#) factor, as well as with DT#. Interestingly, general DT# and originality# were not predicted by openness, but crystallized intelligence (gc#) mediated the relation between openness and DT# ($p = .01$). Extraversion and honesty–humility were weakly associated with DT#. Originality# was not predicted by any of the cognitive ability and

personality traits. In this respect, the model of Study 2 differs from the model presented in Study 1, as the originality[#] factor had substantial variance and therefore could be related to other variables. Interindividual differences in DT[#] were explained to $R^2 = .40$, and the—limited—originality[#] variance remained entirely unexplained ($R^2 = .04$). Note that the variance of the originality[#] factor in Study 2 was significant ($p = .05$), although its factorial saturation in the large model was still very low ($\omega = .19$).

Discussion

Although creativity is of great importance, the progress achieved in understanding creative ability as a construct has been rather limited. In the present studies, we contribute to the answering of two research questions that aim to gain a better understanding of creativity. The first question asks how are individual differences in DT including indicators of fluency and originality structured? Second, how is DT along with originality related with established cognitive abilities, personality traits and insight? In the next sections, we summarize and interpret our findings regarding the two aspects of DT. We will proceed by discussing the relation of DT with convergent thinking, personality, and insight and provide desiderata for further research.

On the dimensionality of divergent thinking

DT is frequently applied to measure creativity, but only very few studies focus on the aspects of DT. Investigating the dimensionality of DT, we provide results on the extent to which originality tasks have residual communalities after an overarching DT factor is controlled for. In Studies 1 and 2, we administered multiple DT tasks including fluency and originality assessments. All measures in both studies were not restricted to the verbal domain: convergent and divergent thinking were assessed in the verbal and figural domain, respectively (Razumnikova, Volf, & Tarasova, 2009). Regarding the dimensionality of DT, competing measurement models favoured a structure including a specific factor of originality[#] besides a general factor of DT[#] (Model D). In the following, we will refer to DT[#] and originality[#] as only DT and originality, without the pound as superscript (#). Model D captures specific variance of originality tasks in a nested originality factor after controlling for general communalities between all fluency and originality tasks. A model with a single DT factor (Model A) fitted the data worse than Model D. We have chosen the nested factor model (Model D) because this model best displays methodological considerations and the theory behind DT as a single construct, rather than being composed of two correlated, but more independent factors (Model B) or being

equally important subcomponents of a higher order factor (Model C).

The fluency factor. In our studies, we show that fluency indicators load high on a general DT factor that has satisfactory reliability (e.g. Dumas & Dunbar, 2014; Forthmann, Jendryczko, et al., 2019). This finding illustrates the important and crucial role of fluency in DT and replicates previous studies that have stressed the importance of fluency also in originality (e.g. Hocevar, 1979). Theoretically, fluency—the quantity of ideas—is a necessary precondition for providing a unique answer (originality). Even though fluency plays an important role in DT and hence in originality, it is wrong to infer that simple fluency tasks measure anything beyond the quantity of ideas. In sum, our overarching DT factor captures the commonality across many diverse tasks including broad retrieval fluency, figural fluency, and verbal fluency. One asset of fluency is its simple measurement (e.g. Batey et al., 2009) and its strong interrater reliability in the relatively easy count of solutions. Fluency tasks provide a time-efficient method for capturing individual differences in the number of generated answers. They are a crucial—but not the only—part of DT.

The originality factor. An important question is whether or not originality can be seen as a distinct latent factor that captures significant variance beyond and above general DT. Originality was considered by numerous prior studies, although its measurement and association with other dimensions of DT and further cognitive constructs are still underexplored and insufficiently understood (e.g. Forthmann, Jendryczko, et al., 2019; Jauk et al., 2014). Nevertheless, from a theoretical perspective, originality is stressed as being more essential than other aspects of DT (e.g. Acar et al., 2017). However, the psychometric properties of the latent variable originality captured above DT from our studies give rise to more questions rather than simply providing a clear answer about the nature of originality. The reliability of the specific originality factor was very low in both of our studies. Please note that a lack of systematic variation between participants as a result of individual differences in originality is only an issue if originality is modelled as a latent ability over and above DT. The specificity of originality as a dimension of individual differences was hampered by the restricted number of tasks we could apply in the present studies. In addition, we aimed to only capture originality with tasks belonging to the verbal content domain.

Previous studies reported low reliability estimates of originality tasks in general (not being controlled for individual differences in fluency) and also evidenced nonsignificant factor variances (Forthmann, Jendryczko, et al., 2019). Nevertheless, with a larger

sample size in Study 2, we were able to show that a specific originality factor had substantial variance advocating its inclusion as a distinct dimension in a comprehensive assessment of DT (see also Acar et al., 2017). Taken together, the two studies do not provide a clear picture on whether or not originality is a better approximation of creativity. Our findings encourage further research to strengthen the measurement of originality. One such approach might be the application of computerized scoring approaches to originality tasks, such as *latent semantic analysis*, in order to assess originality in a more reliable and cost-efficient way (Dumas & Dunbar, 2014). Previous research shows that latent semantic analysis approximates human ratings in evaluating category membership (Laham, 1997) or essay scorings (Foltz, Streeter, & Lochbaum, 2013). Despite its robustness and utility in creativity research (Prabhakaran et al., 2014), further investigation is required for comparisons with human raters and its relation with relevant criteria. At the same time, however, the present findings lead to further questions of setting objective answer standards. Setting such standards (e.g. where is the boundary between a creative and a nonsense answer) is difficult with computerized scorings; moreover, evaluating new scoring methods is mostly based on its comparison with human ratings.

Divergent and convergent thinking

Historically, DT was discussed as a lower level factor of intelligence. In order to assess the uniqueness of DT, we aimed to embed it into a nomological net and examined its relation with a broad variety of cognitive abilities and insight. We tested for relations with general intelligence, crystallized intelligence, and mental speed based on a model that includes measures for fluid intelligence, working memory, mental speed, and different content domains of crystallized intelligence. This makes our studies unique as compared with previous research that is usually based on a more restricted range of cognitive ability indicators. In our first study, we evaluated the relationship between general intelligence and DT. We showed that DT is moderately related with general intelligence and crystallized intelligence. In contrast to previous studies that reported a link between DT and mental speed (e.g. Benedek et al., 2014; Forthmann, Jendryczko, et al., 2019), we found no substantial association. The specific latent factor of originality that explains variance beyond DT was unrelated to any of the investigated cognitive abilities. In our second study, we again found that DT was predicted by general and crystallized intelligence; likewise, DT was unrelated to mental speed. The magnitude of the relationships between DT and general intelligence and crystallized intelligence were slightly higher than relations previously reported in the literature (Kim, 2008). The nonsignificant relationships between originality

and cognitive abilities were contrary to our expectations. We argue that this might be because of the psychometric shortcomings of originality described earlier.

The nomological net provided in Study 2 was also extended by adding insight in order to demonstrate its relations with DT and cognitive abilities. Our results in Study 2 show that the correlation between DT and insight is of similar magnitude as the correlation between DT and cognitive abilities. This implies that insight is not only a variant of intelligence but is also meaningfully correlated with DT. The convergent nature of insight tasks has the advantage to rule out scoring problems and potentially limited interrater reliability associated with other DT tasks. Although, we wish to emphasize that our studies did not focus on typical insight tasks with only one correct answer, but applied anagrams and scrabble tasks in a fluency and originality condition. Insight tasks, such as items from the Remote Associates Test (Mednick, 1962, 1968), are commonly scored as dichotomous variables for correctness only, but further approaches that focus on scoring the originality of such answers have also been proposed (e.g. based on latent semantic analysis; Beisemann, Forthmann, Burkner, & Holling, 2019) and should be considered in the future. In sum, our results indicate that insight is equally correlated with cognitive abilities and DT. Further research is needed to replicate and extend the present results by using a yet broader variety of insight measures.

Divergent thinking and personality traits

In addition to cognitive abilities, we also studied the relationship between DT and personality traits. Based on the literature and to reduce model complexity, we only included personality traits that were previously related with DT (honesty–humility, openness, and extraversion; Silvia et al., 2011). Interestingly, only extraversion and honesty–humility significantly predicted DT. The expected positive relationships between openness and DT, as well as between openness and originality, were not significant. However, crystallized intelligence mediated the relation between openness and DT. That implies that openness is unrelated to DT once crystallized intelligence is controlled for. The relationship between divergent thinking (fluency) and openness and openness and crystallized intelligence has been examined previously (DeYoung, 2015; Kackenmester, Bott, & Wacker, 2019; Kandler et al., 2016; Schretlen, van der Hulst, Pearlson, & Gordon, 2010). Although such studies often report that openness predicts both DT and crystallized intelligence (e.g. a strong relation between openness and fluency and a smaller relation between openness and fluency; Schretlen et al., 2010), studies that have investigated a possible interaction between the three constructs are sparse. Silvia (2008b) reported that openness accounts for the relation

between a *g*-factor (including fluid and crystallized intelligence) and a latent creativity factor, a finding that provides a first hint for the interplay between these constructs. Because of its far-ranging theoretical implications, the reported mediations effects require replications in future studies. As a limitation, our assessment of personality in Study 2 was restricted to the level of overarching factors. Therefore, more fine-grained distinctions (e.g. fantasy versus ideas as facets of openness; Jauk et al., 2014) could not be studied. These distinctions might paint a more detailed picture of the relationships between personality traits, DT, and even crystallized intelligence.

Limitations of the studies

Although our studies included sufficiently large sample sizes and a variety of indicators, there are still limitations that have to be noted regarding the creativity measurement models. Despite the fact that both studies are based on numerous fluency and originality indicators, the number of tasks deployed is imbalanced. Both studies included only one indicator for figural fluency but several verbal indicators. However, a distinction between different content domains was not an objective of these studies. Additionally, in both studies, only two originality indicators but four fluency indicators were used. Although it would be labour intensive for future studies to run additional originality tasks, such studies would probably find substantial variability for a latent originality factor if they did. However, it is uncertain that a stronger originality factor, for example in terms of broader measurements, would result in higher specific variance and show meaningful correlations with covariates.

As in any study, the task selection can be debated. Tasks were selected based among tests validated with German samples (see method section), which led to the inclusion of fluency tasks that are very similar to other prominent DT tasks, for example, the Alternate Uses Task. Apparently, equating psychological constructs with individual tasks is a bad practice. Instead, the multivariate approach pursued here is better suited to represent highly general psychological constructs. We recommend that future studies also include validated and newly devised creativity tasks, which would allow for a better generalization across different creativity measures. In our Study 2, we included insight tasks that were based on anagram and scrabble paradigms. Participants were instructed to respond either fluently or originally. The application of these tasks was somewhat explorative, and their future validation is desirable.

Desiderata for further research

Fortunately, a number of multivariate studies have recently been published. Many of these studies provide strong contributions to the understanding of creativity. In our studies, we aimed to further strengthen these contributions by incorporating a broad set of indicators for measuring creativity, cognitive abilities, personality traits, and insight. Our nested factor model with several fluency and originality tasks extends previous confirmatory modelling of DT (e.g. Dumas & Dunbar, 2014; Silvia, 2008a) and allows us to relate a specific factor of originality with other variables of interest. Specific originality in our studies was not significantly related to any other construct of interest. Such relations have been reported in previous studies, for example weak relations between originality and art grades in school (Forthmann, Jendryczko, et al., 2019) that suffered from psychometric problems like in the present studies (low factor saturation). Future studies might want to further investigate the relations between DT, crystallized intelligence, and openness, as described earlier. The exact nature of this nomological net is likely to be important for tailoring intervention studies.

Previously, research investigates in validity criteria of creativity. Despite systematic research on for example creative outcomes, studies show that they are not necessarily predicted by originality and fluency (e.g. creative achievements; Jauk et al., 2014). Therefore, the dignity of outcomes such as creative achievements might need further validation itself. Besides the extensions of the nomological net mentioned earlier, our understanding of the nature of creativity might also be furthered by studying long-term storage and retrieval (McGrew, 2009) and its relation with creativity (Silvia et al., 2013). Future studies should elaborate on this relation by implementing multiple tasks and confirmatory factor analysis.

We applied different fluency and originality tasks that were only instructed for the construct of interest. Despite our approaches, the nature of originality and its relation with other construct remain unclear. Previous studies have often only applied measures that were coded for both fluency and originality at the same time (Dumas & Dunbar, 2014; Jauk et al., 2014; Silvia, 2008a). This leads to statistical and experimental dependencies that bias the results. We recommend that the study of originality should be based on a variety of DT tests that are only instructed for the construct of interest.

Besides, we assume that a larger and more diverse set of originality tests such as plot titles (Berger & Guilford, 1969) or consequences test (Christensen, Merrifield, & Guilford, 1958) might overcome the encountered reliability issues. Although these tasks are over half a century old, the development of new DT tasks or other creativity measures rarely moves beyond these old assessments. However, there are new tasks that take into account the dynamic nature of the creative process (especially studied in the neuroscience of creativity), such as the *Multi-Trial Creative Ideation* framework (Barbot, 2018).

It assesses fluency by modelling the response time when generating a response, whereas taking into account time for exploration and production (Barbot, 2018). More generally, research on DT needs new approaches and standards (see also Barbot et al., 2019). In particular, originality tasks usually require time-consuming human ratings that often lack sufficient reliability. Although, future studies might profit from applying and evaluating a variety of different scoring approaches (Benedek et al., 2013; Silvia et al., 2008), but the scoring of maximal effort measures is ultimately about delivering a psychometrically sound procedure that evaluates the degree to which participants have succeeded in performing what they were asked to. Therefore, it seems quite promising to investigate in alternative scoring approaches, such as computerized scoring (Forthmann, Oyebade, et al., 2019). Further research is needed to investigate the meaning of such computerized scores.

Conclusion

Central questions about the internal structure and the construct validity of creativity remain unsolved. We have summarized the current state of affairs, applied a comprehensive battery of DT tasks, and compared competing measurement models. We showed that a nested factor model including an overarching DT and a specific originality factor provided good fit to the data. Including both constructs into a nomological net, we found a moderate relationship between intelligence and DT. Insight was correlated with intelligence as well as with DT. Extraversion and honesty–humility predicted DT, whereas crystallized intelligence meditated the relationship between openness and DT. The specific originality factor was neither related with intelligence nor related with personality factors. In sum, fluency appears to be a psychometrically sound construct of the quantity of ideas but is lacking an evaluation of idea quality. Originality as a specific factor, even though of great theoretical importance, shows limited specificity above and beyond DT. We suggest that DT—as measured in our studies—is more than just intelligence, insight, and/or personality. However, in order to better understand DT and originality, further investigations regarding its measurement, modelling, and relationship with relevant outcomes remain essential.

Data accessibility statement

This article earned Open Data and Open Materials badges through Open Practices Disclosure from the Center for Open Science: https://osf.io/tvyxz/wiki. The data and materials are permanently and openly accessible at https://osf.io/c8j29/.

Notes

1. According to a Darwinian theory (Campbell, 1960; Simonton, 1999), the creative process includes two mental mechanisms:blind variation andselective retention. More recent work argues that the theoretical link between these Darwinian mental processes and creativity is problematic (Gabora, 2011).
2. The paper 'Caught in the act: Predicting cheating in unproctored knowledge assessment' (Steger, Schroeders, & Wilhelm, 2020) includes several knowledge measures (e.g. crystallized intelligence) and parts of the personality measure (the honesty–humility scale) from Study 2. The second paper 'It's more about what you do not know than what you know: Testing Competing Claims About Overclaiming' (Goecke, Weiss,Steger, Schroeders, & Wilhelm, 2020) includes several knowledge measures (e.g. crystallized intelligence) and parts of the personality measure (honesty-humility and openness scale) and DT from Study 2. This paper was accepted for publication inIntelligence.

Supplemental material

Supplemental material for this article is available online.

References

Abraham, A. (2018). *The neuroscience of creativity*. University Press. https://doi.org/10.1017/9781316816981

Acar, S., Burnett, C., & Cabra, J. F. (2017). Ingredients of creativity: Originality and more. *Creativity Research Journal*, *29*, 133–144. https://doi.org/10.1080/10400419.2017.1302776

Acar, S., & Sen, S. (2013). A multilevel meta-analysis of the relationship between creativity and schizotypy. *Psychology of Aesthetics, Creativity, and the Arts*, *7*, 214–228. https://doi.org/10.1037/a0031975

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*, 227–257. https://doi.org/10.1016/S0160-2896(96)90016-1

Ackerman, P. L. (2009). On weaving personality into a tapestry of traits. *British Journal of Psychology*, *100*, 249–252. https://doi. org/10.1348/000712609X413737

ACT. (n.d.). *OECD Selects ACT to Develop PISA 2021 Creative Thinking Assessment*. https://connecme.act

club.org/2018/09/oecd-selects-act-to-develop-pisa-2021-creative-thinking-assessment/

Amabile, M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*, 997–1013. https://doi.org/10.1037/0022-3514.43.5.997

Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality*, *15*, 327–353. https://doi.org/10.1002/per.417

Ashton, M. C., & Lee, K. (2009). The HEXAC0-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*, 340–345. https://doi.org/10.1080/00223890902935878

Ball, L. J., Marsh, J. E., Litchfield, D., Cook, R. L., & Booth, N. (2015). When distraction helps: Evidence that concurrent articulation and irrelevant speech can facilitate insight problem solving. *Thinking & Reasoning*, *21*, 76–96. https://doi.org/10.1080/13546783.2014.934399

Barbot, B. (2018). The dynamics of creative ideation: Introducing a new assessment paradigm. *Frontiers in Psychology*, 9. https://doi.org/10.3389/fpsyg.2018.02529

Barbot, B., Hass, R. W., & Reiter-Palmon, R. (2019). Creativity assessment in psychological research: (re)setting the standards. *Psychology of Aesthetics, Creativity, and the Arts*, *13*, 233–240. https://doi.org/10.1037/aca0000233

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

Barron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology*, *32*, 439–476. https://doi.org/10.1146/annurev.ps.32.020181.002255

Batey, M. (2012). The measurement of creativity: From definitional consensus to the introduction of a new heuristic framework. *Creativity Research Journal*, *24*, 55–65. https://doi.org/10.1080/10400419.2012.649181

Batey, M., Chamorro-Premuzic, T., & Furnham, A. (2009). Intelligence and personality as predictors of divergent thinking: The role of general, fluid and crystallised intelligence. *Thinking Skills and Creativity*, *4*, 60–69. https://doi.org/10.1016Zj.tsc.2009.01.002

Batey, M., & Furnham, A. (2006). Creativity, intelligence, and personality: A critical review of the scattered literature. *Genetic, Social, and General Psychology Monographs*, *132*, 355–429. https://doi.org/10.3200/M0N0.132.4355-430

Beier, M. E., & Ackerman, P. L. (2001). Current-events knowledge in adults: An investigation of age, intelligence, and nonability determinants. *Psychology and Aging*, *16*, 615–628. https://doi.org/10.1037//0882-7974.16.4.615

Beisemann, M., Forthmann, B., Burkner, P. C., & Holling, H. (2019). Psychometric evaluation of an alternate scoring for the remote associates test. *The Journal of Creative Behavior.*. https://doi.org/10.1002/jocb.394

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence*, *46*, 73–83. https://doi.org/10.1016/j.intell.2014.05.007

Benedek, M., Könen, T., & Neubauer, A. C. (2012). Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *6*, 273–281. https://doi.org/10.1037/a0027059

Benedek, M., Muhlmann, C., Jauk, E., & Neubauer, A. C. (2013). Assessment of divergent thinking by means of the subjective top-scoring method: Effects of the number of top-ideas and time-on-task on reliability and validity. *Psychology of Aesthetics, Creativity, and the Arts*, *7*, 341–349. https://doi.org/10.1037/a0033644

Berger, R. M., & Guilford, J. P. (1969). *Plot titles*. Sheridan Psychological Services.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*, 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, *67*, 380–400. https://doi.org/10.1037/h0040373

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi. org/10.1017/CB09780511571312

Christensen, P. R., Merrifield, P. R., & Guilford, J. P. (1958). *Consequences: Manual for administration, scoring, and interpretation*. Beverly Hills, CA: Sheridan Supply.

Chu, Y., & MacGregor, J. N. (2011). Human performance on insight problem solving: A review. *The Journal of Problem Solving*, *3*, 119–150. https://doi.org/10.7771/1932-6246.1094

Cropley, A. (1967). Divergent thinking and science specialists. *Nature*, *215*, 671–672. https://doi.org/10.1038/215671a0

Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal*, *18*, 391–404. https://doi.org/10.1207/s15326934crj1803_13

Danthiir, V., Wilhelm, O., Schulze, R., & Roberts, R. D. (2005). Factor structure and validity of paper-and-pencil measures of mental speed: Evidence for a higher-order model?. *Intelligence*, *33*, 491–514. https://doi.org/10.1016/j.intell.2005.03.003

DeCaro, M.S. (2018). When does higher working memory capacity help or hinder insight problem solving? In F. Vallee-Tourangeau (Ed.), *Insight: On the origins of new ideas* (1st ed., pp. 79–104). Routledge https://doi.org/10.4324/97813152 68118-5.

DeYoung, C. G. (2015). Openness/intellect: A dimension of personality reflecting cognitive exploration. In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology, volume 4: Personality processes and individual differences* (pp. 369–399). American Psychological Association https://doi.org/10.1037/14343-017

DeYoung, C. G., Flanders, J. L., & Peterson, J. B. (2008). Cognitive abilities involved in insight problem solving: An individual differences model. *Creativity Research Journal*, *20*, 278–290. https://doi.org/10.1080/10400410802278719

Diedrich, J., Jauk, E., Silvia, P. J., Gredlein, J. M., Neubauer, A. C., & Benedek, M. (2018). Assessment of real-life creativity: The inventory of creative activities and achievements (ICAA). *Psychology of Aesthetics,*

Creativity, and the Arts, *12*, 304–316. https://doi.org/10.1037/aca0000137

Dumas, D., & Dunbar, K. N. (2014). Understanding fluency and originality: A latent variable perspective. *Thinking Skills and Creativity*, *14*, 56–67. https://doi.org/10.1016Zj.tsc.2014.09.003

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*, 541–562. https://doi.org/10.1037/met0000083

Eid, M., Krumm, S., Koch, T., & Schulze, J. (2018). Bifactor models for predicting criteria by general and specific factors: Problems of nonidentifiability and alternative solutions. *Journal of Intelligence*, *6*, 42–65. https://doi.org/10.3390/jintelligence 6030042

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Educational Testing Service.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Eysenck, H. J. (1993). Creativity and personality: Suggestions for a theory. *Psychological Inquiry*, *4*, 147–178. https://doi.org/10.1207/s15327965pli0403_1

Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review*, *2*, 290–309. https://doi.org/10.1207/s15327957pspr0204_5

Foltz, P. W., Streeter, L. A., & Lochbaum, E. (2013). Implementation and applications of the intelligent essay assessor. In M. D. Shermis, & J. Burstein (Eds.), *Handbook of automated essay evaluation* (pp. 90–110). Routledge.

Forthmann, B., Holling, H., Çelik, P., Storme, M., & Lubart, T. (2017). Typing speed as a confounding variable and the measurement of quality in divergent thinking. *Creativity Research Journal*, *29*, 257–269. https://doi.org/10.1080/10400419.2017.1360059

Forthmann, B., Jendryczko, D., Scharfen, J., Kleinkorres, R., Benedek, M., & Holling, H. (2019). Creative ideation, broad retrieval ability, and processing speed: A confirmatory study of nested cognitive abilities. *Intelligence*, *75*, 59–72. https://doi. org/10.1016/j.intell.2019.04.006

Forthmann, B., Oyebade, O., Ojo, A., Gunther, F., & Holling, H. (2019). Application of latent semantic analysis to divergent thinking is biased by elaboration. *The Journal of Creative Behavior*, *53*, 559–575. https://doi.org/10.1002/jocb.240

Forthmann, B., Szardenings, C., & Holling, H. (2018). Understanding the confounding effect of fluency in divergent thinking scores: Revisiting average scores to quantify artifactual correlation. *Psychology of Aesthetics, Creativity, and the Arts*, *14*, 94–112. https://doi.org/10.1037/aca0000196

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for kit of reference tests for cognitive factors (revised 1963)*. Educational Testing Service.

Gabora, L. (2011). An analysis of the blind variation and selective retention theory of creativity. *Creativity Research Journal*, *23*, 155–165. https://doi.org/10.1080/10400419.2011.571187

Gajda, A., Karwowski, M., & Beghetto, R. A. (2017). Creativity and academic achievement: A meta-analysis.

Journal of Educational Psychology, *109*, 269–299. https://doi.org/10.1037/edu0000133

Goecke, B., Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020). Testing Competing Claims About Overclaiming. *Intelligence*, *81*, 101470. https://doi.org/10.1016/j.intell.2020.101470

Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2012). Neural correlates of creativity in analogical reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 264–272. https://doi.org/10.1037/a0025764

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*, 444–454. https://doi.org/10.1037/h0063487

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, *53*, 267–293. https://doi.org/10.1037/h0040755

Guilford, J. P. (1966). Intelligence: 1965 model. *American Psychologist*, *21*, 20–26. https://doi.org/10.1037/h 0023296

Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.

Harrington, D. M. (1975). Effects of explicit instructions to "be creative" on the psychological meaning of divergent thinking test scores. *Journal of Personality*, *43*, 434–454. https://doi.org/10.1111/j.1467-6494.1975.tb00715.x

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, *26*, 433–443. https://doi.org/10.1177/0956797614567339

Hocevar, D. (1979). Ideational fluency as a confounding factor in the measurement of originality. *Journal of Educational Psychology*, *71*, 191–196. https://doi.org/10.1037/0022-0663.7L2.191

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. https://doi.org/10.1080/1070551 99095401187

Jäger, A. O. (2006). *Berliner intelligenzstruktur-testfur jugendliche begabungs-und hochbegabungsdiagnostik; BIS-HB [Berlin structure-of-intelligence test for youth: Diagnosis of talents and giftedness]*. Hogrefe.

Jäger, A. O., Süß, H. M., & Beauducel, A. (1997). *Berliner intelligenzstruktur-test: BIS-test [Berlin structure-of-intelligence test: BIS-test]*. Hogrefe.

Jauk, E., Benedek, M., & Neubauer, A. C. (2014). The road to creative achievement: A latent variable model of ability and personality predictors: The road to creative achievement. *European Journal of Personality*, *28*, 95–105. https://doi.org/10.1002/per.1941

Jung, R. E., Wertz, C. J., Meadows, C. A., Ryman, S. G., Vakhtin, A. A., & Flores, R. A. (2015). Quantity yields quality when it comes to creativity: A brain and behavioral test of the equal-odds rule. *Frontiers in Psychology*, 6. https://doi.org/10.3389/fpsyg.2015.00864

Käckenmester, W., Bott, A., & Wacker, J. (2019). Openness to experience predicts dopamine effects on divergent thinking. *Personality Neuroscience*, *2*, 1–7. https://doi.org/10.1017/pen.2019.3

Kandler, C., Riemann, R., Angleitner, A., Spinath, F. M., Borkenau, P., & Penke, L. (2016). The nature of

creativity: The roles of genetic factors, personality traits, cognitive abilities, and environmental sources. *Journal of Personality and Social Psychology*, *111*, 230–249. https://doi.org/10.1037/pspp0000087

Kaufman, J. C., Baer, J., Cropley, D. H., Reiter-Palmon, R., & Sinnett, S. (2013). Furious activity vs. understanding: How much expertise is needed to evaluate creative work? *Psychology of Aesthetics, Creativity, and the Arts*, *7*, 332–340. https://doi.org/10.1037/a0034809

Kaufman, J. C., & Plucker, J. A. (2011). Intelligence and creativity. In R. J. Sternberg, & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 771–783). Cambridge University Press. https://doi.org/10.1017/CB09780511977244.039

Kim, K. H. (2005). Can only intelligent people be creative? A metaanalysis. *Journal of Secondary Gifted Education*, *16*, 57–66. https://doi.org/10.4219/jsge-2005-473

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance tests of creative thinking (TTCT). *Creativity Research Journal*, *18*, 3–14. https://doi.org/10.1207/s15326934crj1801_2

Kim, K. H. (2008). Meta-analyses of the relationship of creative achievement to both IQ and divergent thinking test scores. *The Journal of Creative Behavior*, *42*, 106–130. https://doi.org/10.1002/j.2162-6057.2008.tb01290.x

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Koehler, W. (1967). Gestalt psychology. *Psychological Research*, *31*, 18–30. https://doi.org/10.1007/BF00422382

Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology*, *65*, 71–93. https://doi.org/10.1146/annurev-psych-010213-115154

Laham, D. (1997). Latent semantic analysis approaches to categorization. *Proceedings of the 19th Annual Conference of the Cognitive Science Society*.

Landmann, N., Kuhn, M., Piosczyk, H., Feige, B., Riemann, D., & Nissen, C. (2014). Entwicklung von 130 deutschsprachigen compound remote associate (CRA)—Wortratseln zur untersuchung kreativer prozesse im Deutschen sprachraum [Development of 130 German compound remote associate (CRA)—Word puzzles for the evaluation of creative processes in the German language area]. *Psychologische Rundschau*, *65*, 200–211. https://doi.org/10.1026/0033-3042/a000223

Lubart, T. I. (2001). Models of the creative process: Past, present and future. *Creativity Research Journal*, *13*, 295–308. https://doi.org/10.1207/S15326934CRJ1334_07

Lubart, T. I., Pacteau, C., Jacquet, A.-Y., & Caroff, X. (2010). Children's creative potential: An empirical study of measurement issues. *Learning and Individual Differences*, *20*, 388–392. https://doi.org/10.1016/j.lindif.2010.02.006

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185–199. https://doi.org/10.1037/0033-2909.114.L185

Martindale, C. (1999). Biological bases of creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 137–152). Cambridge University Press.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology*, *52*, 1258–1265. https://doi.org/10.1037/0022-3514.52.6.1258

McCrae, R. R., & Costa, P. T. (1989). More reasons to adopt the five-factor model. *American Psychologist*, *44*, 451–452. https://doi.org/10.1037/0003-066X.44.2.451

McDonald, R. P. (1999). *Test theory: A unified treatment*. L. Erlbaum Associates.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*, 437–455. https://doi.org/10.1037/a0028085

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220–232. https://doi.org/10.1037/h0048850

Mednick, S. (1968). The remote associates test. *The Journal of Creative Behavior*, *2*, 213–214. https://doi.org/10.1002/j.2162-6057.1968.tb00104.x

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorenstruktur, psychometrische eigenschaften und messinvarianz der Deutschsprachigen version des 60-item HEXACO personlichkeitsinventars. *Diagnostica*, *60*, 86–97. https://doi.org/10.1026/0012-1924/a000112

Mourgues, C. V., Preiss, D. D., & Grigorenko, E. L. (2014). Reading skills, creativity, and insight: Exploring the connections. *The Spanish Journal of Psychology*, 17. https://doi.org/10.1017/sjp.2014.59

Mulaik, S. A., & Quartetti, D. A. (1997). First order or higher order general factor? *Structural Equation Modeling: A Multidisciplinary Journal*, *4*, 193–211. https://doi.org/10.1080/10705519709540071

Mumford, M. D. (2003). Where have we been, where are we going? Taking stock in creativity research. *Creativity Research Journal*, *15*, 107–120. https://doi.org/10.1080/10400419.2003.9651403

Novick, L. R., & Sherman, S. J. (2003). On the nature of insight solutions: Evidence from skill differences in anagram solution. *The Quarterly Journal of Experimental Psychology Section A*, *56*, 351–382. https://doi.org/10.1080/02724980244000288

Nusbaum, E. C., Silvia, P. J., & Beaty, R. E. (2014). Ready, set, create: What instructing people to "be creative" reveals about the meaning and mechanisms of divergent thinking. *Psychology of Aesthetics, Creativity, and the Arts*, *8*, 423–432. https://doi.org/10.1037/a0036549

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, *51*, 45–100. https://doi.org/10.1016/S0079-7421(09)51002-X

Patrick, C. (1935). Creative thought in poets. *Archives of Psychology*, *26*, 1–74.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in python. *Journal of Neuroscience Methods*, *162*, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Prabhakaran, R., Green, A. E., & Gray, J. R. (2014). Thin slices of creativity: Using single-word utterances to assess creative cognition. *Behavior Research Methods*, *46*, 641–659. https://doi.org/10.3758/s13428-013-0401-7

Preckel, F., Holling, H., & Wiese, M. (2006). Relationship of intelligence and creativity in gifted and non-gifted students: An investigation of threshold theory.

Personality and Individual Differences, *40*, 159–170. https://doi.org/10.1016Zj.paid.2005.06.022

Preckel, F., Wermer, C., & Spinath, F. M. (2011). The inter-relationship between speeded and unspeeded divergent thinking and reasoning, and the role of mental speed. *Intelligence*, *39*, 378–388. https://doi.org/10.1016/j.intell.2011.06.007

Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relationships of personality to differential conceptions of creativity: A systematic review. *Psychology of Aesthetics, Creativity, and the Arts*, *11*, 59–68. https://doi.org/10.1037/aca0000079

PWC. (2016). *19th annual global CEO survey*. http://www.pwc.com/gx/en/ceo-survey/2016/landing-page/pwc-19th-annual-global-ceo-survey.pdf

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge https://doi.org/10.4324/9780203841624.

Razumnikova, O. M., Volf, N. V., & Tarasova, I. V. (2009). Strategy and results: Sex differences in electrographic correlates of verbal and figural creativity. *Human Physiology*, *35*, 285–294. https://doi.org/10.1134/S0362119709030049

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*, 544–559. https://doi.org/10.1080/00223891.2010.496477

Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*, *13*, 144–152. https://doi.org/10.1037/aca0000227

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research*. Northwestern University.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Runco, M. A. (2004). Creativity. *Annual Review of Psychology*, *55*, 657–687. https://doi.org/10.1146/annurev.psych.55.090902.141502

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, *24*, 66–75. https://doi.org/10.1080/10400419.2012.652929

Runco, M. A., & Albert, R. S. (1985). The reliability and validity of ideational originality in the divergent thinking of academically gifted and nongifted children. *Educational and Psychological Measurement*, *45*, 483–501. https://doi.org/10.1177/001316448504500306

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

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53–61. https://doi.org/10.1007/BF02289209

Schmiedek, F., & Li, S.-C. (2004). Toward an alternative representation for disentangling age-associated differences in general and specific cognitive abilities. *Psychology and Aging*, *19*, 40–56. https://doi.org/10.1037/0882-7974.19.L40

Schmitz, F., Rotter, D., & Wilhelm, O. (2018). Scoring alternatives for mental speed tests: Measurement issues and validity for working memory capacity and the attentional blink effect. *Journal of Intelligence*, *6*, 47–77. https://doi.org/10.3390/jintelligence6040047

Schmitz, F., & Wilhelm, O. (2016). Modeling mental speed: Decomposing response time distributions in elementary cognitive tasks and correlations with working memory capacity and fluid intelligence. *Journal of Intelligence*, *4*, 13–36. https://doi.org/10.3390/jintelligence4040013

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan, & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 73–163). The Guilford Press.

Schooler, J. W., & Melcher, J. (1995). The ineffability of insight. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 97–133). The MIT Press.

Schoppe, K. J. (1975). *Verbaler kreativitats-test-VKT: Ein verfahren zur crfassung verbal-produktiver kreativitäts-merkmale [Verbal creativity tests—VKT: A test for measuring verbal-productive creativity characterizes]*. Verlag für Psychologie CJ Hogrefe.

Schretlen, D. J., van der Hulst, E.-J., Pearlson, G. D., & Gordon, B. (2010). A neuropsychological study of personality: Trait openness in relation to intelligence, fluency, and executive functioning. *Journal of Clinical and Experimental Neuropsychology*, *32*, 1068–1073. https://doi.org/10.1080/13803391003689770

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Silvia, P. J. (2008a). Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan. *Creativity Research Journal*, *20*, 34–39. https://doi.org/10.1080/10400410701841807

Silvia, P. J. (2008b). Another look at creativity and intelligence: Exploring higher-order models and probable confounds. *Personality and Individual Differences*, *44*, 1012–1021. https://doi.org/10.1016/j.paid.2007.10.027

Silvia, P. J. (2015). Intelligence and creativity are pretty similar after all. *Educational Psychology Review*, *27*, 599–606. https://doi.org/10.1007/s10648-015-9299-1

Silvia, P. J., Beaty, R. E., & Nusbaum, E. C. (2013). Verbal fluency and creativity: General and specific contributions of broad retrieval ability (Gr) factors to divergent thinking. *Intelligence*, *41*, 328–340. https://doi.org/10.1016/j.intell.2013.05.004

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

Silvia, P. J., Martin, C., & Nusbaum, E. C. (2009). A snapshot of creativity: Evaluating a quick and simple method for assessing divergent thinking. *Thinking Skills and Creativity*, *4*, 79–85. https://doi.org/10.1016/j.tsc.2009.06.005

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., et al. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective

scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 68–85. https://doi.org/10.1037/1931-3896.2.2.68

Simonton, D. K. (1999). Creativity as blind variation and selective retention: Is the creative process Darwinian?. *Psychological Inquiry*, 10, 309–328. https://doi.org/10.1207/S15327965PLI1004_4

Simonton, D. K. (2011). Creativity and discovery as blind variation and selective retention: Multiple-variant definition and blind-sighted integration. *Psychology of Aesthetics, Creativity, and the Arts*, 5, 222–228. https://doi.org/10.1037/a0023144

Sprugnoli, G., Rossi, S., Emmendorfer, A., Rossi, A., Liew, S.-L., Tatti, E., di Lorenzo, G., ... Santarnecchi, E. (2017). Neural correlates of Eureka moment. *Intelligence*, 62, 99–118. https://doi.org/10.1016/j.intell.2017.03.004

Steger, D., Schroeders, U., & Wilhelm, O. (2019). On the dimensionality of crystallized intelligence: A smartphone-based assessment. *Intelligence*, 72, 76–85. https://doi.org/10.1016/_j.intell.2018.12.002

Steger, D., Schroeders, U., & Wilhelm, O. (2020). Caught in the act: Predicting cheating in unproctored knowledge assessment. *Assessment*. Advance online publication. https://doi.org/10.1177/10731911209149701073191120 91497.

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

Sternberg, R. J., & Lubart, T. I. (1993). Investing in creativity. *Psychological Inquiry*, 4, 229–232. https://doi.org/10.1207/s15327965pli0403_16

Sternberg, R. J., & O'Hara, L. A. (1999). Creativity and intelligence. In R. Sternberg (Ed.), *Handbook ofcreativity*. Cambridge University Press.

Süß, H. M., & Beauducel, A. (2005). Faceted models of intelligence. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence*. Sage Publications https://doi.org/10.4135/9781452233529.n18.

Taylor, C. L. (2017). Creativity and mood disorder: A systematic review and meta-analysis. *Perspectives on Psychological Science*, 12, 1040–1076. https://doi.org/10.1177/1745691617699653

Vartanian, O., Martindale, C., & Kwiatkowski, J. (2007). Creative potential, attention, and speed of information processing. *Personality and Individual Differences*, 43, 1470–1480. https://doi.org/10.1016/j.paid.2007.04.027

Vernon, P. E. (1971). Effects of administration and scoring on divergent thinking tests. *British Journal of Educational Psychology*, 41, 245–257. https://doi.org/10.1111/j.2044-8279.1971.tb00669.x

Vock, M., Preckel, F., & Holling, H. (2011). Mental abilities and school achievement: A test of a mediation hypothesis. *Intelligence*, 39, 357–369. https://doi.org/10.1016/j.intell.2011.06.006

von Stumm, S., & Ackerman, P. L. (2013). Investment and intellect: A review and meta-analysis. *Psychological Bulletin*, 139, 841–869. https://doi.org/10.1037/a0030746

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it?. *Frontiers in Psychology*, 4. https://doi.org/10.3389/fpsyg.2013.00433

Wilhelm, O., & Schroeders, U. (2019). Intelligence. In R. Sternberg, & J. Funke (Eds.), *The psychology of human thought: An introduction* (pp. 257–277). Heidelberg University Publishing.

Wilhelm, O., Schroeders, U., & Schipolowski, S. (2014). *Berliner test zur erfassung fluider und kristalliner intelligenz fur die 8. Bis 10. jahrgangsstufe [Berlin test of fluid and crystallized intelligence for grades 8–10]*. Hogrefe.

Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). The measurement of individual differences in originality. *Psychological Bulletin*, 50, 362–370. https://doi.org/10.1037/h0060857