Towards a Better Understanding of Individual Differences in Creativity and Improving its Measurement

Kumulative Dissertation

zur Erlangung des Doktorgrades Dr. rer. nat. der Fakultät für Ingenieurwissenschaften, Informatik und Psychologie der Universität Ulm

Vorgelegt von:

Selina Jutta Klara Weiß

aus Nürnberg

2020

Institut für Psychologie und Pädagogik

Abteilung Differentielle Psychologie und Psychologische Diagnostik

Leitung: Prof. Dr. Oliver Wilhelm

Tag der Promotion: 18. März 2021

Amtierender Dekan: Prof. Dr.-Ing. Maurits Ortmanns

1. Gutachter: Prof. Dr. Oliver Wilhelm
2. Gutachter: Prof. Dr. Aljoscha Neubauer
3. Gutachter: Dr. Patrick Kyllonen

Acknowledgments

Writing my dissertation was a journey full of very different experiences. It was interesting, joyful, educational, sometimes a little painful, and for sure laborious. However, it was worth every second. I believe I can say I have learned and grown a lot through these years, and I would do it again immediately. Especially because I have had the best supervisor, I could have asked for. Oliver Wilhelm, thank you – I guess there are not many words I need to say: You really rock! Without you, it would not have been as much fun, as educational, and as encouraging.

However, many other people joined, shared, supported, and supervised my journey. First of all, I want to thank Benjamin Goecke for his endless support, love, and patience. You having my back guided me through the last years. Besides, you are a great researcher, and talking to you always expanded my horizon. I would never have wanted to miss traveling, working, living, and loving (with) you! But, there are a lot of other fantastic junior researchers that have shared all the joy (and pain) with me: Johanna Hartung, we have been so much more than office co-workers throughout all the years, and I can still remember our first data collection in Berlin as it was yesterday. Diana Steger, the evenings with wine and long conversations were much more than just generating new ideas for research. Mattis Geiger, what would have been all these years without beer and BBQ at your place? The list of colleagues that I had the best time I could have asked for is long: Thank you, Sally Olderbak, Gabriel Olaru, Florian Schmitz, Luc Watrin, and many more. I am also truly thankful for all my senior co-authors who have provided their knowledge and their valuable feedback: Ulrich Schroeders, you have taught me a lot, and your advice and the chitchat was always great. Andrea Hildebrandt, apart from your professional advice, it was a pleasure to have a female professor to talk to. Patrick Kyllonen, your literature knowledge was a treasure.

The most important thing to say is that the manuscripts I present in this dissertation would not have been possible without my brilliant colleagues and co-authors' contributions! In short, it was my pleasure, and at this point, I also want to thank everyone else who made my contribution to science possible (numerous research assistants, Annette Kniep).

Lastly, I want to thank my family. Petra and Peter Weiss, you have provided me with the freedom to choose my way in life and have educated me in a way that made me curious, open, creative, and self-confident enough to start that journey. I also want to thank my sisters, Chiara and Pamina Weiss. You have always supported me, and both your professions provided new angles on my research regarding creativity.

Table of Content

VI. Epilogue

VII. Appendix

List of Tables

List of Figures

List of Manuscripts

The manuscripts presented in this dissertation are published in or submitted for publication to the following journals:

I. Weiss, S., & Wilhelm, O. (2020). Coda: Creativity in psychological research versus in linguistics – Same but different? *Cognitive Semiotics*, *13*(1), 1-15. https://doi.org/10.1515/cogsem-2020-2029

II. Weiss, S., Wilhelm, O., & Kyllonen, P. (under review, submitted 2020). A review and taxonomy of creativity measures. *Psychology of Aesthetics, Creativity, and the Arts*.

III. Weiss, S., Steger, D., Kaur, Y., Hildebrandt, A., Schroeders, U., & Wilhelm, O. (2020). On the trail of creativity: Dimensionality of divergent thinking and its relation with cognitive abilities and personality. *European Journal of Personality, 0*(0), 1-22. https://doi.org/10.1002/per.2288

IV. Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020). A Reappraisal of the Threshold Hypothesis of Creativity and Intelligence. *Journal of Intelligence*, *8*(4), 38-58. https://doi.org/10.3390/jintelligence8040038

Abstract

Individual differences in creativity have been studied and measured for over a century. However, our understanding of creativity as a cognitive ability and its relation to other psychological constructs is still limited. This might partly be due to a) lack of interdisciplinary collaboration, b) low convergent validity of measures that supposedly assess creativity, and c) the small number of multivariate studies investigating the nomological net of creativity. This dissertation addresses these limitations by studying the interdisciplinary understanding of creativity (Manuscript I), its measurement (Manuscript II), and the nomological net of creativity (Manuscript III and IV).

The first Manuscript considers and discusses an interdisciplinary view on similarities and dissimilarities between creativity in applied linguistics and psychology. Based on this review, I introduce and relate common terminology and discuss domain-general processes in both disciplines. I conclude by stating that applied linguistics can contribute important expertise for improving the psychometrics of performance appraisals through creativity measures in psychology.

In the second Manuscript, I develop a taxonomy for categorizing creativity indicators and summarize the creativity measurements applied since the 1900s based on an exhaustive literature review. This taxonomy—including measurement approaches, definitions of constructs, data types, scoring, and psychometric problems—allows assessing the convergent validity of measures and the development of new and more innovative ones. It is concluded that the application of divergent thinking tasks (tapping aspects such as fluency and originality) shows the smallest discrepancies between the standard definition of creativity as ability and the measures' scope.

In the third Manuscript, a selection of these tasks was included in two extensive multivariate studies to shed new light on measurement models of creativity and embed creativity into a nomological net of established ability constructs, personality, and insight. The

findings imply that a distinction between the two major aspects of divergent thinking (fluency and originality) is psychometrically difficult. Additionally, the results encourage an interpretation of divergent thinking as an ability construct that is more than just a linear combination of intelligence, personality, and insight.

In a fourth study, the threshold theory is thoroughly investigated. It states that an individual's creativity might benefit from intelligence at low intelligence levels, but not once intelligence is above a certain threshold. Based on hitherto proposed and improved methodological approaches and two data sets, I conclude that intelligence and creativity are linearly related across the continuum of intelligence.

In total, these four Manuscripts provide guidance in understanding and measuring creativity as a unique psychological construct. This inspires future research approaches, such as a) new assessments, b) alternative scoring procedures, and c) an enlargement of the nomological net of creativity by potentially related constructs (e.g., retrieval ability and emotional creativity).

I. Prologue

"The quest for easily objectifiable testing and scoring has directed us away from the attempt to measure some of the most precious qualities of individuals and hence ignore those qualities."

- Joy P. Guilford (1950)

Creativity has often been described as one of the key aspects of being successful in life (Gabora & Kaufman, 2010). After decades of creativity research, it became a pressing topic again as new technologies for measurement and data analysis have become available, and various workforce and educational programs have started shifting their focus towards creativity in the last years. Arguably, creativity is the ability that enables us to accomplish essential contributions in a rather computerized world, where any computer or artificial intelligence is more efficient, smarter, and faster than us (Bughin et al., 2018, May 23). Creativity might be the only cognitive ability that still can supplement a machine-driven world (PWC, 2016, January). The relevance of being creative leads to a high prevalence of this explicit area in various research branches. For example, studies incorporating neuroimaging to explore creativity's neurological basis apply innovative measurements and technologies (e.g., Kaur et al., 2020). As another example, the educational sector focuses on how children can be educated to be more creative (e.g., Berg et al., 2012).

However, the importance of creativity has been recognized long before the *fourth industrial revolution*, resulting in numerous studies and scientific discussion about creativity in the last century. Despite its great importance and a long history of research, this *skill* (Guilford, 1950) or *trait* (Eysenck, 1993) is still somewhat sparsely understood when it comes to the most basic questions: What is creativity? How can we measure it? And what are its relations with other important constructs, such as general intelligence, retrieval ability, or openness to experience? This dissertation sheds light on some of these questions as it aims to further the understanding of creativity and tidy up in the jungle that arose from the use of different terminologies, definitions, and measurements over decades. In order to do so, this dissertation presents four manuscripts that theoretically and empirically tackle these questions. Before presenting these manuscripts, I provide a general theoretical overview regarding the definition of creativity, its research history, its measurement, and its relation to neighboring constructs.

The Research History of Creativity

In the middle of the twentieth century, Joy P. Guilford leveled the path for decades of creativity research in psychology to come (Guilford, 1950). His attempts to understand creativity and related processes (Guilford, 1956), individual differences in creativity (Wilson et al., 1953), and the measurement of creativity (Wilson et al., 1954) were and still are pioneering. For example, Wilson and colleagues (1954) provided factor analytical studies of fifty-three tests that measure creative-thinking and extracted factors that were widely used later on (e.g., various fluency factors, flexibility factors, and originality). However, creativity has also been studied before Guilford, although with a different focus and different approaches. Earlier studies and investigations mostly focused on originality. People's originality has been reviewed before the 1950s primarily based on single-case studies of eminent geniuses (Cattell, 1903). Cattell (1903), for example, rated 1000 men listed in at least three dictionaries with the most extensive descriptions (e.g., Napoleon, Shakespeare, Voltaire, Luther), regarding their originality.

Nevertheless, studies of creativity were rare before the second half of the twentieth century, and journals solely devoted to creativity research developed much later (see Sternberg & Lubart, 1999). After Guilford brought creativity to the attention of a whole research community (1950), creativity has been studied based on different theoretical considerations and from different angles—some to be taken more seriously than others (Sternberg & Lubart, 1999). Some attempts had a somewhat mystical understanding of creativity (e.g., Muses as a source of inspiration; Ghiselin, 1985) or a psychodynamic understanding of it (e.g., creative expression as a form of unconscious wishes; Freud, 1910). Studies that mainly contribute to nowadays understanding of creativity are focusing either on understanding creativity as a cognitive-skill or an ability (e.g., Guilford, 1950), as a trait or a confluence of several personality traits (e.g., Eysenck, 1993), or as both, a combination of personality traits, motivation, and different types of cognitive abilities (e.g., Amabile, 1983; Mumford & Gustafson, 1988). Even though the latter—often referred to as confluence theory (Sternberg & Lubart, 1999)—is represented by various researchers, it should be scrutinized, for example, with regard to the existence of multiple types of intelligences. These different research branches can also be seen in the varieties of developed measures (Weiss et al., under review, submitted 2020). However, this diversity of how creativity is studied and understood might have its roots in its definitional vagueness.

Defining Creativity

The definition of creativity goes back to the nineteenth century. Bethune (1839) defined a genius as someone who can produce "new combinations of thoughts," and he stressed that "originality of conception" is an essential part of being a genius (Bethune, 1839, p. 59). Since then, various creativity definitions were published and proposed (e.g., Runco & Jaeger, 2012). Reviewing some of these definitions leads to the conclusion that most of them have a bipartite intention in common that goes back to Barron (1955). Barron stressed that an original idea must be useful in reality and therefore clearly separable from random and meaningless responses. Many definitions include these two parts: that creativity is, on the one hand, based on a truly novel and original idea, but on the other hand, this idea is useless if it is not appropriate (e.g., Mumford, 2003; Runco & Jaeger, 2012). This definition is widely recognized (Runco & Jaeger, 2012). Despite its acceptance, it provides room for a lot of interpretation. It is quite vague. This leads to diversity in research and myths due to the large range of possible interpretations (e.g., Plucker et al., 2004; Simonton, 2012).

Besides, the definition has been troubled by a) different disciplines requiring different interpretations of novelty and appropriateness, b) approaches of domain-specificity, c) the epoch and Zeitgeist, and finally, d) the human who is judging the originality and appropriateness. First, creativity is not only studied in psychology but also within fields like linguistics. In the first manuscript, I relate the above-described definition to the terminology that has been used in linguistics for describing creativity in the language (Weiss & Wilhelm, 2020). One example of a different interpretation is what is considered novel in psychology versus linguistics, given that language has its norms that restrict the range of new word interventions. Next, domain-specificity approaches have stated that creativity might not be an overarching construct and requires different definitions and measurement approaches. Several domain-specific approaches (e.g., Diedrich et al., 2018; Kaufman, 2012) root their domains in Greek mythology, where nine muses inspire creative outcome in different areas, such as poesy (e.g., lyrics), but also astrology or history (D'Aulaire & D'Aulaire, 1992; Kaufman, 2012). These domain-specific approaches have been driven as far as describing 56 domains (Kaufman et al., 2009). This diversity challenges a general standard definition of creativity. The question whether such domain richness is warranted and theoretically acceptable is also discussed in Weiss and Wilhelm (2020). Third, the epoch and Zeitgeist restricts what is considered creative. For example, the classical composer Claude Debussy was an unrecognized genius for a long time. He neglected the Eurocentric Zeitgeist and included free rhythmical elements inspired by Indonesian folk music (Handschik, 2015). From the end of the nineteenth until the mid of the twentieth century, the Zeitgeist favored creative inventions based on a strict rhythm corset and the tonality chosen, for example, by Richard Wagner. Therefore, Debussy's compositions were highly novel, but at that time, they were often seen as inappropriate. This leads to the last point that troubles the definition of creativity: the human judge. Mostly, creativity is evaluated and judged by humans. They are not only affected by Zeitgeist, but they also have different levels of expertise in judging the value of an invention or different tastes when it comes to various types of artistic performances. For example, defining creativity by originality and appropriateness might appear generally helpful for judges. But when it comes to areas such as art or any kind of artistic performance the value or appropriateness is difficult to account for and might be highly subjective. In science the actual value of a creative technological achievement might be easier to assess, while judging the originality of an invention can be difficult.

In sum, a vague and very broad standard definition of creativity exists, but that does not solve several problems and comes along with unique challenges. Despite all efforts, the understanding of creativity is still limited. As theories and the conceptual understanding are vague—ranging from typical behavior to maximum effort approaches of creativity—the measurement of creativity is even more challenging. In the next section, I will elaborate on previous measurement attempts and scoring procedures of creativity.

The Measurement of Creativity

As described above, decades of research have failed to develop a straightforward standard definition of creativity, and it gets vaguer when it comes to the measurement of creativity. On the one hand, it is discussed if creativity can be measured at all (Eysenck, 1994; Piffer, 2012). On the other hand, many measures have been published that arguably all capture creativity – or not (Runco, 2008). The literature includes an extensive range of measures, often using varying labels (see Weiss et al., under review, submitted 2020). In the following introduction on creativity measurement, I focus on an introductory description of maximal effort and typical behavior (e.g., Eysenck, 1993; Guilford, 1950). A more detailed and fine-grained overview in terms of a taxonomy that enables a specific allocation of measures is provided in the second manuscript (also including example measures and limitations, that will not be further discussed here; see Weiss et al., under review, submitted 2020). The differentiation between maximal effort and typical behavior is dated back to Cronbach (1949) and is found in many constructs, like personality and intelligence.

Measures of Typical Behavior

Typical behavior describes how a person behaves, feels, or believes under ordinary circumstances. This usually implies that someone is asked about their everyday feelings, actions, or favorites and disfavors. Therefore, self-reports are mostly applied as a measurement of typical behavior, but also others can be asked about a person's typical believes and behaviors. In the second manuscript, I describe the ratings of such preferences or accomplishments in detail (Weiss et al., under review, submitted 2020). The application of such assessments of typical behavior—for example, to measure creative attitudes—is prevalent as the measurement and scoring appear straightforward (Forgeard & Kaufman, 2016; Kaufman et al., 2008). However, the assessment of typical behavior comes with a list of disadvantages, such as validity restrictions and individuals' introspection (Kaufman, 2019; Silvia, 2008). I argue that the psychometric and theoretical problems (e.g., faking, biases, convergent validity, ability and willingness to introspect, or evaluation of originality/novelty) outweigh the possible advantages.

Measures of Maximal Effort

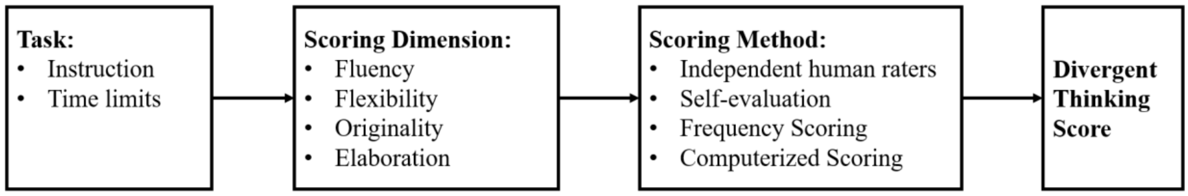
In contrast to the measurement approach mentioned above, it is argued that creativity might be better captured if a creative performance is measured. Performance measures are fundamentally different from measures of typical behavior. Measures of maximal performance are usually administered in standardized settings in a laboratory, and the participant is aware and willing to show a specific performance (Wilhelm & Schroeders, 2019). These measures are objects to the scientific discussions, questioning if maximal effort measures can be applied to assess creativity (Runco, 2008). Another critical question is how maximal effort can be scored as a clear veridical answer is absent due to the nature of such tasks (Reiter-Palmon et al., 2019). In the following sections, I describe the measurement and scoring of maximal effort creativity in more detail. To do so, a prominent class of maximal effort measures in creativity research, namely the divergent thinking measures, are introduced (for more information, see Weiss et al., under review, submitted 2020).

Divergent Thinking Measures. In the *Structure-of-Intellect* model, Guilford described divergent thinking (divergent production) as the cognitive processes necessary for creativity (Guilford, 1956). Later on, divergent tasks have become the most prominent measure of creativity in the literature, although they have suffered a loss of popularity in the last years (Said-Metwaly et al., 2017; Snyder et al., 2019). Divergent thinking tasks have some advantages over typical behavior measures (for more information and example items, see Weiss et al., under review, submitted 2020). For example, they provide a standardized measure of a performance that can be evaluated, and they are known for better psychometric properties in the context of creativity research than assessments of typical behavior (Runco & Acar, 2012). One critique of divergent thinking tests is that they seem to be no guarantor for life's actual behavior (Runco, 2008; Runco & Acar, 2012). Another challenge that these tasks are related to is their scoring (e.g., Cseh & Jeffries, 2019). In the following, I will elaborate on that and describe different scoring approaches and challenges.

Scoring Divergent Thinking. First, the instruction (e.g., "*be creative"*) is crucial and leads to variance in the given answers (Harrington, 1975; Nusbaum et al., 2014). Second, the scoring dimension is critical and should be in line with the instruction. Compared to maximal effort measures with one veridical solution that serves as a benchmark for the behavior (Wilhelm & Schroeders, 2019), the range of responses to divergent thinking tasks is broad, and multiple solutions are correct, while some are more creative than others. Therefore, divergent thinking can be evaluated by fluency (quantity), flexibility (diversity), originality (the quality of an answer), and elaboration (specificity of a solution; Carroll, 1993; Torrance, 1966). The scoring and instruction of verbal fluency are very common in the literature as the scoring appears the easiest (Silvia et al., 2013). The scoring of flexibility and elaboration is less common in the literature, arguably due to a lack of specific tasks and scoring guidelines. Another widespread approach is the scoring of originality that can follow one of the later named procedures (Reiter-Palmon et al., 2019): a) scoring guidelines for independent human raters taking into account the uncommonness, remoteness, cleverness of an answer (Hocevar, 1979; Wilson et al., 1953); b) self-evaluation of the participants (Silvia et al., 2008), c) semi-automatized frequency scorings (e.g., frequency in the total sample; Vernon, 1971), or d) computerized scorings based on the latent distance of answers (latent semantic analysis; e.g., Forthmann et al., 2018). Figure 1 schematically summarizes all the above-described decisions a researcher has to make for getting one score of divergent thinking (Reiter-Palmon et al., 2019).

Figure 1

Schematic Presentation of Decisions towards a Divergent Thinking Score



Creativity and Related Constructs

The definition (see II. Manuscript I), the measurement of creativity (see III. Manuscript II), and its scoring have been discussed and problematized already. Still, to understand creativity and isolate it from closely related constructs, it is crucial to take its nomological net into account. Cronbach and Meehl (1955) stated that "learning more about a theoretical construct is a matter of elaborating the nomological network in which occurs" (Cronbach & Meehl, 1955, p. 187). Therefore, creativity is described concerning closely related constructs in the following sections. This broadens our understanding of the construct itself and validates measures in terms of convergent and discriminant validity. The focus is on two constructs, namely personality, and intelligence that have been previously related (closely) with creativity and are prominent constructs for explaining individual differences. However, the sections will only serve as a broader introduction for the two empirical manuscripts that provide results regarding the nomological net (see IV. Manuscript III) and investigate the relation between creativity and intelligence (see V. Manuscript IV).

Creativity and Personality

Personality has often been related to creativity, primarily when creativity was understood as a trait (e.g., Eysenck, 1993). Besides, personality arguably explains individual differences in creative performances. This leads to one question: What personality factors have an impact on being creative? Personality is commonly described in terms of five factors (conscientiousness, extraversion, agreeableness, neuroticism, and openness; McCrae & Costa, 1989) that were initially derived in lexical analysis of the language (e.g., Allport & Odbert, 1936). Descriptions of dimensions were derived based on adjectives from *Webster's Unabridged Dictionary of English Language* (Allport & Odbert, 1936; Norman, 1967). Norman (1967) used 75 categories in his taxonomy, including over 1000 adjectives that described a dimension. For example, 17 adjectives (such as insightful, clever, curious, and creative) were later summarized under the openness/intellect factor in the facet originality (Goldberg, 1990; Norman, 1967). The debate about the number of personality dimensions—for example, three dimensions sensu Eysenck (1991), five dimensions sensu McCrae and Costa (1989), or six dimensions sensu Ashton and Lee (2001)—is beyond the scope of this dissertation. Therefore I focus on the most common approaches of five and six dimensions (i.e., Big Five and

HEXACO). Both models have been related to creativity repeatedly, resulting in a broad range of relations. What can be summarized is that most previous studies, based on the five and six personality dimensions, reported a significant correlation of creativity with openness (e.g., Batey & Furnham, 2006; Feist, 1998; McCrae, 1987; Puryear et al., 2017; Silvia et al., 2011). For a more detailed description of the relations' magnitude, see the third manuscript (Weiss et al., 2020a). However, studies that focus on the larger nomological net and include intelligence for a broader picture of such correlations are sparse. To better understand the nomological net, the relation between creativity and intelligence is described in the next section.

Creativity and Intelligence

Historically, creativity and intelligence were always related to another, as creativity was embedded in various intelligence models. Providing an overview of intelligence models is beyond the scope of this dissertation. Some prominent representatives are described in the following (for an overview, see Wilhelm & Schroeders, 2019). Spearman (1904) introduced the *g*-factor view on intelligence, implying that one general factor captures all tasks' correlations and indicates test-specific individual differences. In contrast to that view is the idea that multiple correlated factors display intelligence (e.g., seven primary factors of intelligence; Thurstone, 1938). These early attempts have led to accepted models in intelligence research; for example, *Carroll's Three-Stratum Theory* described three generality levels (Carroll, 1993). The model is based on the analysis of 461 data sets and is one of the models that include creativity in terms of a retrieval factor that captures ideational fluency and word fluency, and sensitivity to problems. This implies that creativity was considered a lower-level factor of general intelligence (*g*). Another model that embedded creativity underneath general intelligence is the *Berlin Intelligence Structure Model* (Jäger et al., 1997). It covered "inventiveness" (divergent thinking) within three traditionally considered content domains (figural, verbal, and numerical). As already described above, the earliest attempt to embed divergent thinking into the structure of human cognitive abilities was proposed by Guilford's (1967) *Structure-of-Intellect Model*. Guilford stressed the contribution of divergent thinking for creativity and convergent thinking (ability to identify one correct solution). In the second half of the twentieth century, there were still doubts about the distinction of creativity and intelligence concerning their hierarchical integration (e.g., Getzels & Jackson, 1962; Wallach & Kogan, 1965). This hierarchy and the association between intelligence and divergent thinking have been discussed and emphasized ever since (e.g., Benedek et al., 2012; Preckel et al., 2011; Runco, 2004; Silvia et al., 2013). Not only the question to what magnitude creativity and intelligence are related (see IV. Manuscript III for an overview; Weiss et al., 2020a) are subject to numerous studies, but also the question if they are related linearly (see V. Manuscript IV; Weiss et al., 2020b) throughout all levels of intelligence (Guilford & Christensen, 1973; Guilford, 1967; Jauk et al., 2013; Karwowski et al., 2016; Preckel et al., 2011). Despite numerous studies, there are quite some questions unanswered, while other answers persist in the literature, even though they might not be justified.

Overview of the Dissertation Manuscripts

One key to furthering our knowledge regarding creativity is to comprehend the construct by learning how creativity can be assessed. Another important part is establishing a nomological net that includes the above presented close constructs. Studies that include such essential factors might provide several answers and lead to new interesting research questions. In the next sections, I present four manuscripts along these lines. Based on the manuscripts presented below, I provide an in-depth view of creativity in an interdisciplinary context, measuring creativity and uncovering creativity's nomological net. The first manuscript (*II. Manuscript I: Creativity in Psychological Research versus in Linguistics – Same same but different?*) provides a theoretical overview that integrates the theories and definitions derived in the psychological literature and connects them to the considerations driven in linguistics. This review sets the standards for common terminology and discusses issues beyond the definition, such as measurement and scoring. The second manuscript (*III. Manuscript II: A Review and Taxonomy of Creativity Measures*) includes another critical issue: The creativity measurement. This review summarizes a large variety of measures and provides a taxonomy that enables a distinct categorization of such. The third manuscript (*IV. Manuscript III: On the Trail of Creativity: Dimensionality of Divergent Thinking and its Relation with Cognitive Abilities, Personality, and Insight*) builds upon the first and second manuscript and brings the definition as well as the measurement towards an empirical work. In two multivariate studies, a large task selection is analyzed to shed light on the structure of divergent thinking and the nomological net of it by including ability, personality, and insight. In a fourth manuscript (*V. Manuscript IV: A Reappraisal of the Threshold Hypothesis of Creativity and Intelligence*), the relation between intelligence and creativity is thoroughly discussed and analyzed based on two large studies and three different analytical approaches. This manuscript adds to the discussion of a threshold's existence and how such a threshold can be approached.

References

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1), 1-171. https://doi.org/10.1037/h0093360

Amabile, T. M. (1983). *The social psychology of creativity.* Springer-Verlag.

Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality*, *15*(5), 327–353. https://doi.org/10.1002/per.417

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

Batey, M., & Furnham, A. (2006). Creativity, intelligence, and personality: A critical review of the scattered literature. *Genetic, Social, and General Psychology Monographs*, *132*(4), 355–429. https://doi.org/10.3200/MONO.132.4.355-430

Benedek, M., Franz, F., Heene, M., & Neubauer, A. C. (2012). Differential effects of cognitive inhibition and intelligence on creativity. *Personality and Individual Differences*, *53*(4), 480–485. https://doi.org/10.1016/j.paid.2012.04.014

Berg, H., Taatila, V., & Volkmann, C. (2012). Fostering creativity – a holistic framework for teaching creativity. *Development and Learning in Organizations: An International Journal*, *26*(6), 5–8. https://doi.org/10.1108/14777281211272242

Bethune, G. W. (1839). Genius. *Casket*, *8*, 59–69.

Bughin, J., Hazan, E., Lund, S., Dahlström, P., Wiesinger, A., & Subramaniam, A. (2018, May 23). *Skill shift: Automation and the future of the workforce*. https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. University Press.

Cattell, J. M. (1903). A statistical study of eminent men. *Popular Science Monthly*, *62*, 359–377.

Cronbach L. J. (1949). *Essentials of psychological testing.* Haper.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Cseh, G. M., & Jeffries, K. K. (2019). A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 159–166. https://doi.org/10.1037/aca0000220

D'Aulaire, I., & D'Aulaire, E. P. (1992). *D'Aulaires' book of Greek myths*. Delacorte.

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

Eysenck, H. J. (1991). Dimensions of personality: 16, 5, or 3? Criteria for a taxonomic paradigm. *Personality and Individual Differences*, *12*(8), 773–790. https://doi.org/10.1016/0191-8869(91)90144-Z

Eysenck, H. J. (1993). Creativity and personality: Suggestions for a theory. *Psychological Inquiry*, *4*(3), 147–178. https://doi.org/10.1207/s15327965pli0403_1

Eysenck, H. J. (1994). The measurement of creativity. In M. A. Boden (Ed.), *Dimensions of creativity* (pp. 199–242). The MIT Press.

Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review*, *2*(4), 290–309. https://doi.org/10.1207/s15327957pspr0204_5

Forgeard, M. J. C., & Kaufman, J. C. (2016). Who cares about imagination, creativity, and innovation, and why? A review. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(3), 250–269. https://doi.org/10.1037/aca0000042

Forthmann, B., Oyebade, O., Ojo, A., Günther, F., & Holling, H. (2018). Application of Latent Semantic Analysis to Divergent Thinking is Biased by Elaboration. *The Journal of Creative Behavior*, *53*(4), 559-575. https://doi.org/10.1002/jocb.240

Freud, S. (1910). Eine Kindheitserinnerung des Leonardo Da Vinci. *Schriften Zur Angewandten Seelenkunde*, *7*, 1–71.

Gabora, L., & Kaufman, S. B. (2010). Evolutionary approaches to creativity. In J. C. Kaufman & R. Sternberg (Eds.), *The Cambridge Handbook of Creativity* (pp. 279–300). Cambridge University Press.

Getzels, J. W., & Jackson, P. W. (1962). *Creativity and intelligence: Explorations with gifted students.* Wiley.

Ghiselin, B. (1985). *The creative process: Reflections on the invention in the arts and sciences*. University of California Press.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

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*(4), 267–293. https://doi.org/10.1037/h0040755

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

Guilford, J. P., & Christensen, P. R. (1973). The one-way relation between creative potential and IQ*. *The Journal of Creative Behavior, 7*(4), 247–252. https://doi.org/10.1002/j.2162-6057.1973.tb01096.x

Handschik, M. (2015). *Musik als "Medium der sich selbst erfahrenden Wahrnehmung": Möglichkeiten der Vermittlung Neuer Musik unter dem Aspekt der Auflösung und Reflexion von Gestalthaftigkeit*. Georg Olms Verlag.

Harrington, D. M. (1975). Effects of explicit instructions to "be creative" on the psychological meaning of divergent thinking test scores. *Journal of Personality*, *43*(3), 434–454. https://doi.org/10.1111/j.1467-6494.1975.tb00715.x

Hocevar, D. (1979). Ideational fluency as a confounding factor in the measurement of originality. *Journal of Educational Psychology*, *71*(2), 191–196. http://dx.doi.org/10.1037/0022-0663.71.2.191

Jäger, A. O., Süß, H. M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: BIS-Test*. Hogrefe.

Jauk, E., Benedek, M., Dunst, B., & Neubauer, A. C. (2013). The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence, 41*(4), 212–221. https://doi.org/10.1016/j.intell.2013.03.003

Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D. M., Gajda, A., Chruszczewski, M. H., & Benedek, M. (2016). Is creativity without intelligence possible? A Necessary Condition Analysis. *Intelligence, 57*(3), 105–117. https://doi.org/10.1016/j.intell.2016.04.006

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. John Wiley & Sons.

Kaufman, J. C., Cole, J. C., & Baer, J. (2009). The construct of creativity: Structural model for self-reported creativity ratings. *The Journal of Creative Behavior*, *43*(2), 119–134. https://doi.org/10.1002/j.2162-6057.2009.tb01310.x

Kaufman, J. C. (2012). Counting the muses: Development of the Kaufman Domains of Creativity Scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*, *6*(4), 298–308. https://doi.org/10.1037/a0029751

Kaufman, J. C. (2019). Self-assessments of creativity: Not ideal, but better than you think. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 187–192. https://doi.org/10.1037/aca0000217

Kaur, Y., Ouyang, G., Sommer, W., Weiss, S., Zhou, C., & Hildebrandt, A. (2020). What does temporal brain signal complexity reveal about verbal creativity? *Frontiers in Behavioral Neuroscience*, *14*(146), 1-17. https://doi.org/10.3389/fnbeh.2020.00146

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology*, *52*(6), 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*(2), 451–452. https://doi.org/10.1037/0003-066X.44.2.451

Mumford, M. D. (2003). Where have we been, where are we going? Taking stock in creativity research. *Creativity Research Journal*, *15*(2-3), 107–120. https://doi.org/10.1080/10400419.2003.9651403

Mumford, M. D., & Gustafson, S. B. (1988). Creative syndrome: Integration, application, and innovation. *Psychological Bulletin*, *103*(1), 27–43. https://doi.org/10.1037/0033-2909.103.1.27

Norman, W. T. (1967). *2800 Personality trait descriptors: Normative Operating Characteristics for a University Population*. Michigan University.

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*(4), 423–432. https://doi.org/10.1037/a0036549

Piffer, D. (2012). Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research. *Thinking Skills and Creativity*, *7*(3), 258–264. https://doi.org/10.1016/j.tsc.2012.04.009

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, *39*(2), 83–96. https://doi.org/10.1207/s15326985ep3902_1

Preckel, F., Wermer, C., & Spinath, F. M. (2011). The interrelationship between speeded and unspeeded divergent thinking and reasoning, and the role of mental speed. *Intelligence, 39*(5), 378–388. https://doi.org/10.1016/j.intell.2011.06.007

Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relating Personality and Creativity: Considering What and How We Measure. *The Journal of Creative Behavior, 53*(2), 232-245. https://doi.org/10.1002/jocb.174

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

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*(2), 144–152. https://doi.org/10.1037/aca0000227

Runco, M. A. (2004). Creativity. *Annual Review of Psychology*, *55*(1), 657–687. https://doi.org/10.1146/annurev.psych.55.090902.141502

Runco, M. A. (2008). Commentary: Divergent thinking is not synonymous with creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 93–96. https://doi.org/10.1037/1931-3896.2.2.93

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, *24*(1), 66–75. https://doi.org/10.1080/10400419.2012.652929

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

Said-Metwaly, S., Noortgate, W. V. den, & Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. *Creativity. Theories – Research - Applications*, *4*(2), 238–275. https://doi.org/10.1515/ctra-2017-0013

Silvia, P. J. (2008). Discernment and creativity: How well can people identify their most creative ideas? *Psychology of Aesthetics, Creativity, and the Arts*, *2*(3), 139–146. https://doi.org/10.1037/1931-3896.2.3.139

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (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*(2), 68–85. https://doi.org/10.1037/1931-3896.2.2.68

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

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*(5), 328–340. https://doi.org/10.1016/j.intell.2013.05.004

Simonton, D. K. (2012). Taking the U.S. patent office criteria seriously: A quantitative three-criterion creativity definition and its implications. *Creativity Research Journal*, *24*(2-3), 97–106. https://doi.org/10.1080/10400419.2012.676974

Snyder, H. T., Hammond, J. A., Grohman, M. G., & Katz-Buonincontro, J. (2019). Creativity measurement in undergraduate students from 1984–2013: A systematic review. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 133–143. https://doi.org/10.1037/aca0000228

Spearman, C. (1904). "General Intelligence" objectively determined and measured. *American Journal of Psychology*, *15*, 201–293. https://doi.org/doi:10.2307/1412107

Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In R. J. Sternberg (Ed.), *Handbook of Creativity*. Cambridge University Press.

Thurstone, L. L. (1938). *Primary mental abilities*. University of Chicago Press.

Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking—Norms, Technical Manual Research Edition—Verbal Tests, Forms A and B—Figural Tests, Forms A and B.* Personnel Press.

Vernon, P. E. (1971). Effects of administration and scoring on divergent thinking tests. *British Journal of Educational Psychology*, *41*(3), 245–257. https://doi.org/10.1111/j.2044-8279.1971.tb00669.x

Wallach, M. A., & Kogan, N. (1965). A new look at the creativity-intelligence distinction1. *Journal of Personality*, *33*(3), 348–369. https://doi.org/10.1111/j.1467-6494.1965.tb01391.x

Weiss, S., Steger, D., Kaur, Y., Hildebrandt, A., Schroeders, U., & Wilhelm, O. (2020a). On the trail of creativity: Dimensionality of divergent thinking and its relation with cognitive abilities and personality. *European Journal of Personality*. https://doi.org/10.1002/per.2288

Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020b). A reappraisal of the threshold hypothesis of creativity and intelligence. *Journal of Intelligence*, *8*(4), 38-58. https://doi.org/10.3390/jintelligence8040038

Weiss, S., Wilhelm, O., & Kyllonen, P. (under review, submitted 2020). A review and taxonomy of creativity measures. *Journal of Aesthetics, Creativity, and the Arts.*

Weiss, S., & Wilhelm, O. (2020). Coda: Creativity in psychological research versus in linguistics – Same but different? *Cognitive Semiotics*, *13*(1), 1-15. https://doi.org/10.1515/cogsem-2020-2029

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

Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). The measurement of individual differences in originality. *Psychological Bulletin*, *50*(5), 362-370. https://doi.org/10.1037/h0060857

Wilson, R. C., Guilford, J. P., Christensen, P. R., & Lewis, D. J. (1954). A factor-analytic study of creative-thinking abilities. *Psychometrika*, *19*(4), 297–311. https://doi.org/10.1007/BF02289230

II. Manuscript I

Coda: Creativity in psychological research versus in linguistics – Same but different?

The literature review and research were done by myself. Concept/Idea, discussion of the results, and the manuscript drafts and the revisions were conducted by myself and supervised and reviewed by Prof. Dr. Oliver Wilhlem. The manuscript was published in Cognitive Semiotics on the 13th of July 2020. Republished with permission of Walter de Gruyter and Company, from Coda: Creativity in psychological research versus in linguistics – Same but different?, Selina Weiss & Oliver Wilhelm, Volume 13, Edition 1, 2020; permission conveyed through Copyright Clearance Center, Inc.

Weiss, S., & Wilhelm, O. (2020). Coda: Creativity in psychological research versus in linguistics – Same but different? *Cognitive Semiotics*, *13*(1), 1-15. https://doi.org/10.1515/cogsem-2020-2029

Selina Weiss* and Oliver Wilhelm

Coda: Creativity in psychological research versus in linguistics – Same but different?

https://doi.org/10.1515/cogsem-2020-2029

Abstract: Understanding the very nature of creativity is a hot topic in research across various disciplines and has profound societal relevance. In this contribution, we discuss verbal creativity by highlighting its definition, psychometric measurement, and relations with other personality dispositions. We relate psychological research with findings from linguistics presented in this issue and depict similarities and differences between both approaches. More specifically, we relate the linguistic terminology of F-creativity to fluency and flexibility, whereas we identify E-creativity as akin to originality. We propose latent semantic analysis as a possible approach for evaluating originality and compare this approach with more commonly applied human ratings. Based on contributions in this issue, we discuss creativity as a domain-general process that is (e. g., in applied arts) often driven by the recombination of mental elements. Lastly, we propose several intelligence and personality dispositions as determinants of individual differences in creativity. We conclude that creativity research in linguistic and psychology has many communalities and interdisciplinary work bears strong promises for the future.

Keywords: verbal creativity, domain-specificity, personality, intelligence

1 Introduction

Creativity is the subject of research across disciplines and has resulted in a wide variety of studies and discussions for over a century. While there are numerous debates regarding the nature of creativity, its definition, its measurement, and its relation with other constructs, one thing on which most researchers agree is that creativity is crucial and essential for being successful in life (Gabora and Kaufman 2010). Hennessey and Amabile (2010) describe the great interest in

***Corresponding author: Selina Weiss**, Ulm University, Ulm, Germany,
E-mail: selina.weiss@uni-ulm.de
Oliver Wilhelm: Ulm University, Ulm, Germany, E-mail: oliver.wilhelm@uni-ulm.de

creativity in flourishing research in this area. This can be also exemplified by a quick and simple PsycINFO search (a comprehensive library of research in psychological science): the search term *Creativity produces over 100,000 articles published between 1859 and 2019, as well as a large amount of publication outlets for creativity. In particular, the majority of articles were published in the 1980s and 1990s. But just now, creativity is in the spotlight again as large-scale initiatives such as the international educational achievement measure PISA recognize the importance of assessing and fostering creativity (OECD: Vincent-Lancrin 2017). This recent interest is derived from several workforce studies and surveys (e. g., PWC 2016) that highlight creativity as the most important human resource in a computerized world. Hence, society and the research community both recognize the significance of defining, understanding, measuring, evaluating, and fostering creativity. Besides, this societal and research interest creativity remains interesting as it is present in everyday life. This general interest is displayed in numerous blogs, books, videos, and magazines that tackle creativity.

The numerous studies often only contribute fragments to the understanding of creativity. Therefore, we are in need of a deeper understanding of creativity through more interdisciplinary research (Hennessey and Amabile 2010).

With our contribution to this special issue, we aim to provide an interdisciplinary overview based on psychological research on creativity. Referring to the other contributions, we elaborate on the definition of creativity, criteria for measurability, the discussion regarding the domain specificity of creativity along with the demands of various domains, and the individual differences in creativity while focusing on verbal creativity.

2 Definition of creativity

Creativity research has a long history and some approaches date back to the early twentieth century (e. g., rank-ordering the most eminent people in history to analyze their creativity; Cattell 1903). Since then, researchers from different disciplines have attempted to define creativity. Varying definitions and applications of creativity emerged as a product of the given discipline and relevant trends at that time. Some definitions focused on problem solving (Polya 1945) or insight (e. g., Gestalt psychology; Wertheimer 1945), while others focused on specific domains (e. g., creativity in poets; Patrick 1935). Despite the variety of different approaches employed, the core of most definitions applied nowadays is bipartite and includes an *original product (new, unusual, novel, unexpected)* that is somehow *valuable* *(useful, good, adaptive, appropriate)* (Batey 2012; Mumford 2003; Runco and Jaeger 2012). Even though the definition is bipartite, both parts are necessary to describe creativity. Based on this definition, a product cannot be creative without being original and appropriate at the same time. This definition also appears throughout this special issue in several varieties: the interaction of *originality/novelty* and *appropriateness* (Hoffmann 2020); the creation of *novel* form-meaning pairs (Turner 2020); the ability to generate *novel* work (e. g., creative instances of language use) that is *appropriate* in the initial context of use (Trousdale 2020); the *lack of formal rigidity* (oral poetics) that results in a recombination and modification of the formulaic system with a greater focus on *flexibility* than on novelty (Págan Cánovas 2020); the production and understanding of *novel* output that is constrained by the computational linguistic system (Bergs and Kompa 2020); *originality* and *appropriateness*, whereas the latter is not leading to novel ideas in language use, but only the violation of the norm can result in creativity regarding language (Uhrig 2020). The small variation in these definitions shows that the appropriateness of a new (word-)invention can be in conflict with the language system in question and might thus be difficult to assess. In contrast to appropriateness, we agree that "valuable" (in the sense of pragmatic utility) might be a better fit terminology-wise (Barron 1955). Although a general consensus has been reached regarding this bipartite definition — save some approaches, such as adding surprise as an important character of a novel product (Simonton 2012), for example — the field still lacks a comprehensive and universal definition and a precise operationalization based on this definition (Parkhurst 1999; Plucker et al. 2004). This shortcoming might be seen as intrinsic to the topic: If an overarching and comprehensive definition could be provided and if established measures of it existed, then the behavior might not constitute creative acts but something else (for example, inductive generalization).

In addition to these general approaches to define creativity by novelty and value, further efforts to describe more fine-grained creative actions and thoughts can be examined.

The research on creativity in psychology was mainly influenced by Guilford 1950 presidential address at the American Psychological Association. He described divergent thinking (or divergent production) as a cognitive process that leads to original and novel outcome. According to Guilford 1950, divergent thinking is a pivotal thought process that results in original and valuable outcomes, although it has to be mentioned that the terms divergent thinking and creativity cannot be used interchangeably (Runco 2008). This ability to think creatively includes four different dimensions: Fluency (quantity of responses), originality (quality of responses), flexibility (variety of responses), and elaboration (number and quality of details provided) (Guilford 1956, 1960). We think that these dimensions provide a nice approximation of creative thought and action. They can be related with a common definition from linguistics that is featured prominently in this special issue (Bergs and Kompa 2020; Hoffmann 2020): the differentiation between *F-creative* (fixed and known possibility space) and *E-creative* (extending or enlarging the existing system) (Sampson 2016). Arguably, F-creativity demonstrates performance in a prescribed space of possible answers. Theoretically, this can be related to fluency and flexibility, as a high performance in both requires the controlled retrieval in a fixed category, as well as switches between given categories (Silvia et al. 2013). Nevertheless, the performance in a flexibility task can lead to an E-creative outcome if the retrieval among existing categories is exceeded. While elaboration can take place in both — a fixed space and in an extended system — originality seems to be theoretically connected to E-creativity. Originality requires not only the retrieval in a given category but also the expansion of the system in order to provide a truly novel and valuable solution. The theoretical link we proposed above requires further empirical validation in an interdisciplinary study. Although F-creative and E-creative are commonly-applied terms in linguistics, it is important to state that true E-creativity might be difficult to conduct in Construction Grammar, as very creative language applications might be considered as wrong because of the given language norm (Uhrig 2020). Therefore, it is questionable whether someone can be original in Construction Grammar if it is only defined by truly new and novel outcomes. Likewise, this reality also holds in most other areas where novel interventions are considered as a once-in-a-lifetime creative achievement rather than something that can be embedded in everyday creative activities. However, Construction Grammar in particular is bound by several rules that can hinder or restrict the production of original outcomes as they might be considered more incorrect than original. Nonetheless, originality is pivotal and merits consideration. In psychological research, originality is mostly assessed by uniqueness (something unique is unexpected and unusual). Therefore, the question about which objective criteria can be applied to measure creativity in language arises.

3 Criteria and measurability of creativity in language

While the definition of creativity is complicated, its operationalization and measurement along with objective criteria of scoring is even more difficult. As stated by Bergs and Kompa 2020, the definition of constraints is crucial to avoid measuring random behavior instead of creativity. In the literature, there are hundreds of tests that can be applied just to measure creativity (Weiss et al. 2020). The literature provides several ways to categorize and describe such measures: a) by defining the object of observation (e. g., the person that is creative, the process of creativity, the environment [press], and the product/outcome [Rhodes 1961]); or b) by adding measurement approaches (e. g., self-ratings) and levels (e. g., teams) to the object of observation (Batey 2012). However, these aforementioned heuristic frameworks are not exhaustive and do not allow for a distinct categorization of measurements. All creativity tests can be subsumed in one of the following categories/measurement approaches: tests that are self-reports (reports of typical/everyday behavior), reports of others (others' of the target's typical behavior), and ability tests (test data regarding verbal and figural production). Weiss and colleagues (forthcoming) provide an overview of creativity measurement and embed prominent measures in a taxonomy. Self-report measures can either include reports of typical/everyday behavior (e. g., the Epstein Creativity Competencies Inventory for Individuals: "I always keep a recording device by my bed at night." [Epstein et al. 2008]) or provide frequencies of creative achievements or actions (e. g., Inventory of Creative Activities and Achievements: "How often have you written a short literary work (e. g., poem, short story) in the past 10 years?" [Diedrich et al. 2018]). Others-report includes others' (like peers, supervisors, experts, etc.) judgment of typical creative behavior (e. g., Scales for Rating the Behavioral Characteristics of Superior Students: "The student demonstrates creative thinking about scientific topics." [Renzulli et al. 2002]) or products, performances, and actions (e. g., the Patent Index [Owens et al. 1957]). When assessing verbal creativity, ability tests are the most helpful in terms of reliable and valid measures. Ability tests mostly include open-ended questions with the requirement to verbally (occasionally figurally) provide a creative idea on a given topic (e. g., "Name as many different ways as you can that you can use a brick!"). A prominent creativity test is the Torrance Test of Creative Thinking (Torrance 1966). This test includes — as most divergent thinking tests do — the actual production of verbal and figural ideas and outcomes. Verbal subtests in the Torrance Test include dimensions such as asking, guessing, and causes while the figural section focuses on picture, construction, and completion. Although the assessment of verbal outcomes or verbally-presented behaviors (as in self-report items) allows for an approximation towards creativity, it remains important to acknowledge that creativity transcends the verbal domain.

Although verbal divergent thinking tests are commonly applied to assess creativity, as in linguistics the question remains what the criterion is to distinguish wrong answers from creative ones: Is "cut down a big tree with a brick" as an answer on alternative uses for a brick wrong or creative? The above presented bipartite definition of creativity presents appropriateness/value as a criterion to distinguish wrong from creative. Nevertheless, the appropriateness/value of a product/thought are problematic to judge as the value itself for example is driven by zeitgeist and culture. Even though variables like zeitgeist and culture might lead to an underestimation of great ideas and products, the gold standard of judging creativity is with human scorings of originality and appropriateness. The literature provides a huge variety of scoring techniques for such tasks (Reiter-Palmon et al. 2019). One frequently-applied technique is the subjective scoring of answers based on human raters (Silvia et al. 2008). An example for such a scoring is the consensual assessment technique, which is based on experts' rating on a five-point scale that resulted from a consensus of these experts (Amabile 1982; Kaufman et al. 2013). This technique has proven and reliable answers, although it comes with disadvantages. For example, experts might consider different definitions of creativity or might understand/interpret one of its definitions differently (e. g., is novelty defined by a statistical rareness or also by remoteness towards other answers?). Moreover, human scorings are time-consuming and lack true, objective criteria for scoring: If one rater would rate "cutting down a big tree" as wrong while another other would rate it as creative, the raters could either discuss the rating or find a consensus or the ratings would just be averaged.

Linguistic systems provide effective standards and criteria for what can be considered as wrong versus right. So why not learn from these systems and try to find a criteria or method that at least can help us improve the objectivity of creativity measurements by serving as an additional standard? A relatively recent objective scoring criteria that seems promising in its application on creativity tasks is the computerized evaluation of open-ended answers based on latent semantic analysis (Landauer et al. 1998). The idea behind these criteria is that creative ideas and hence creative verbal outcomes on these tasks should be remote (Wilson et al. 1953) — remote to other answers or remote to the question. This remoteness can be measured by the semantic similarity/distance of answers (Forthmann et al. 2018). Latent semantic analysis has already proven good reliability in tasks that only require a one-word answer (Prabhakaran et al. 2014). Though, as the answer grows longer, the elaboration or amount of details present could bias the results (Forthmann et al. 2018). Besides, such objective scorings lack an evaluation of the value/appropriateness of an answer or may score unique answers as creative even though they are not creative (Silvia et al. 2008). In sum, this score merits further evaluation in a wider range of tasks and different types of answers (e. g., one word vs. one sentence vs. a short story). Still, we believe that this field and its quite new application on creativity tasks is promising. At this end, the interdisciplinary work of linguists and psychologists can help further elaborate and expand such computerized scoring approaches with the aim of finding more objective criteria to score creative outcomes.

4 Domain-specificity of creativity

As described above, most ability tests that measure creativity focus on the verbal output of a person. Nevertheless, the hallmark of creativity research is the ongoing debate about the dimensionality of creativity itself. This debate includes the question about how domain-general or domain-specific skills and creativity are. A domain general skill or trait would indicate that someone who is creative in one area has a higher probability of also being creative in other areas (Silvia et al. 2009). For example, the domain general view would indicate that a person who shows creative accomplishments in visual arts is also a creative writer. Therefore domain-generality should be understood as similarities in the ability of being creative (or not) in various areas. The general definition of creativity implies that a person has original and valuable output in various areas or has novel and appropriate ideas on a variety of topics. Focusing on an even more fine-grained aspect of creativity another definition would be: A person who shows high fluency, high originality, and high flexibility not only in verbal production, but also in figural production.

Previously, creativity was described as domain-general, including a g-factor view about creativity as a transcending skill or trait (Root-Bernstein 1984). An opposing view discusses creativity as (highly) domain-specific (e. g., domain specificity in educational settings; Barbot et al. 2015 ; e. g., domain specificity in creative accomplishments; Silvia et al. 2009). This diversity highlights an important question: is creativity domain-specific in the way of broad domains (Baer 1998), task-specific (Baer 1993), or domain-general?

In his contribution, Trousdale 2020 examines to what extent musical and linguistic creativity are similar versus different. He finds similarities such as the adaptation of small items within a larger unit (e. g., improvisation) and differences such as the lack of musical semantics. In sum, he concludes that there are general properties in both systems. Moreover, the argumentation of Trousdale 2020 raises the question whether the differences between creativity in music, language, math, science, etc. are minor: Rather, perhaps a general system from that creativity is applied (e. g., recombination of existing patterns). Likewise, Turner 2020 explained that a domain-general process is the basis of creativity in blending. Even in domains that are rather skill-related and require a certain level of expertise (e. g., oral poetics; Págan Cánovas 2020), creativity can be achieved by the recombination and reshaping of existing patterns and chunks — as is the case in jazz improvisation (Trousdale 2020). All these contributions show that different domains have varying features in common yet some that are unique. If these domains share enough communalities, they can all be combined under a general factor of creativity. Psychological literature does not provide multivariate studies large enough to prove or counter any view in a statistically sophisticated way. In sum, we believe that the above-described contributions raise an interesting and relevant point: Is there something that remains as a domain-specific component (besides the skill that is necessary to perform creatively in this domain at all) after controlling for general processes such as reorganization and recombination? Probably not.

What might be more applicable in case of domain-specificity is a focus on who is creative in what context. As presented above, measures of previous creative accomplishments include a variety of domains in which someone can be creative — for example, literature, music, arts and crafts, creative cooking, or even sports. Moreover, these creative achievements can be further divided into rather great accomplishments for mankind (*Big-C*; Kaufman and Beghetto 2009) and accomplishments in areas of everyday-creativity (*little-c*; Richards et al. 1988). They can be even further divided into creative accomplishments in the area of learning and personal experiences (Beghetto and Kaufman 2007; Kaufman and Beghetto 2009). Previous creative accomplishments can be measured in all these different areas and all of them might be applicable in linguistics.

5 Individual differences in verbal creativity

There is no question that being creative in a specific domain is driven by a level of skill in that domain. If that is the case, then what explanation can account for the fact that everyone is not similarly talented in oral poetics? This art form might indeed appear akin to everyday spoken language, yet only few people can successfully perform this specific type of creative expression. Págan Cánovas, 2020 offers a first hint on why some people are able while others are not: differences in working memory capacities and retrieval. These differences are referred to as individual differences and are very important to explain the diversity in creative ability and creative outcome. In the following paragraphs, we elaborate on individual differences that have been previously related to creativity and have already been expressed by Hoffmann 2020: Intelligence (e. g., Kim 2005) and personality (e. g., McCrae 1987).

5.1 Intelligence and working memory

In his contribution to this special issue, Págan Cánovas 2020 describes that oral poetics and the general handling of language in patterns requires chunking and therefore involves working memory capacity and retrieval. The nature of oral poetics thus indicates that creativity in language requires a specific amount of working memory. Working memory capacity reflects the differences between persons regarding the capacity of their "cognitive system responsible for providing access to information required for ongoing cognitive processes" (Wilhelm et al. 2013 *p*.1). Hence, the capacity of this storage might distinguish very creative oral poets from uncreative ones. In the psychological literature, creativity was often linked — even discussed as being the same — to intelligence. In the 80s, there were still doubts about the distinction of creativity and general intelligence (Wallach and Kogan 1965). This discussion goes back to the hierarchical integration of creativity in several prominent intelligence models. Creativity can be understood in the framework of the Structure-of-Intellect (SOI) model (Guilford 1966), the Berliner Intelligenzstrukturmodell (Jäger et al. 1997), and the Three-Stratum Modell (Carroll 1993). Recent findings support such hierarchical integration and report substantial correlations between broad retrieval ability and creative fluency/originality (Silvia et al. 2013). However, the hierarchy between intelligence and creativity has also been scrutinized lately (Preckel et al. 2011). Meta-analytic findings suggest that intelligence and creativity are only weakly related to one another (r = 0.17; 95% CI [0.17, 0.18]; N = 45,880; Kim 2005). Academic achievement and creativity are also only moderately correlated (r = 0.22; 95% CI [0.19, 0.24]; N = 52,578; Gajda et al. 2017). It is important to note that these results might arise due to very different operationalizations and a wide variety in the measurement of creativity (e. g., higher correlations for verbal ability tests with academic achievement [r = 0.30] than for self-reported creativity with academic achievement [r = 0.12]; Gajda et al. 2017). Such theoretical allocations along with a large body of diverse findings might question the relationship between creativity and intelligence, although Silvia and colleagues (2013) state, based on a thorough multivariate study, that the connection between creativity and intelligence might be closer than previously described in the creativity literature. Working memory shows similar results: It only weakly predicts creativity (Benedek et al. 2014). The sum of these findings indicates that working memory and intelligence are important to be creative and that individual differences in creativity go back to differences in working memory (Págan Cánovas 2020) or in knowledge and vocabulary (Bergs and Kompa 2020; Hoffmann 2020; Págan Cánovas 2020; Trousdale 2020; Uhrig 2020). However, it also implies that, in order to be creative, we need more than intelligence (knowledge) or working memory and that these are not the same — even though one might require the other, especially in linguistics. Investigations in the so-called threshold theory have shown that there is arguably a nonlinear relation between intelligence and creativity (Holling and Kuhn 2008). There is a great debate between researchers reporting a threshold for creativity: Intelligence is proposed as a pre-condition for creativity when under this threshold value, whereas neither are as connected when over the threshold (Jauk et al. 2013).

On the other hand, other researchers criticize the various reported thresholds and rather deem intelligence as necessary for creativity (Karwowski et al. 2016).

5.2 Personality

Personality traits, like extraversion and openness, have been connected to language acquisition and blending (Hoffmann 2020; Turner 2020). Being open towards new ideas, values, or fantasies seems to be quite in line with being creative. In the Big Five personality framework, the factor Openness and its underlying traits (such as fantasy, curiosity, and flexibility) are generally linked to creativity (Feist 1998; Furnham and Batey 2006) and especially linked to divergent thinking (verbal creative ability) ($r = 0.39$; McCrae 1987). The literature in the past 30 years has consistently report a small to moderate relation between openness and creativity, while other personality traits (such as conscientiousness, neuroticism, and agreeableness) have mostly been unrelated to creativity (Kandler et al. 2016; McCrae 1987; Puryear et al. 2017). In detail, extraversion has sometimes been found to be positively related to divergent thinking (Martindale and Dailey 1996). The factors conscientiousness, neuroticism, and agreeableness are mostly unrelated to creativity (McCrae 1987), even though some findings for example describe artists as more anxious, emotional, and sensitive, while others show that highly-creative scientists are lower in conscientiousness and agreeableness (Feist 1998). Besides, other personality traits, such as psychoticism (Barron and Harrington 1981) or honesty-humility (Silvia et al. 2011), have been previously related to creativity: For example, participants with high creativity values show lower honesty-humility. In sum, we see again that our personality may influence how prone to creativity we are, but it is clearly distinguishable from creativity itself.

6 General conclusion

In sum, we have showed that linguistics and psychology share similarities when defining, understanding, measuring, and evaluating creativity. Psychological research and its applications highly rely on verbal production and outcomes as a proxy of creativity. While several problems might arise when using this approach, it also enables us to directly link creativity in psychology and linguistics. At this end, interdisciplinary work can provide a better inside in understanding creativity as a whole as well as measuring creativity (Hennessey and Amabile 2010). Such interdisciplinary studies can be driven by a) unifying terminology in order to prevent jangle fallacies (e. g., [preventing] different terms for the same constructs if they are equivalent; Kelley 1927); b) finding empirical support for domain-general processes (e. g., can we find processes in psychology that are domain-general, inspired by conceptual blending, that were described as domain-general and necessary for performing *E-creativity* Hoffmann 2019); and c) elaborating on objective scorings of verbal creativity (e. g., such as computerized scoring approaches as LSA or Construction Grammar). Alongside all the communalities and differences in the aforementioned areas, we can all agree on the importance of the construct for all disciplines. Although the topic is of such a great interest, there is a long way to go before we can claim to understand creativity.

References

Amabile, Teresa M. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology* 43(5). 997–1013.

Baer, John. 1993. *Divergent thinking and creatrivity: A task-specific approach*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Baer, John. 1998. The case for domain specificity of creativity. *Creativity Research Journal* 11(2). 173–177.

Barbot, Baptiste, Besancon, Maud & Lubart Todd. 2015. Creative potential in educational settings: its nature, measure, and nurture. *Education* 43(3). 371–381.

Barron, Frank & David M. Harrington. 1981. Creativity, intelligence, and personality. *Annual Review of Psychology* 32(1). 439–476.

Barron, Frank. 1955. The disposition toward originality. *The Journal of Abnormal and Social Psychology* 51(3). 478–485.

Batey, Mark. 2012. The measurement of creativity: From definitional consensus to the introduction of a new Heuristic framework. *Creativity Research Journal* 24(1). 55–65.

Beghetto, Ronald A. & James C. Kaufman. 2007. Toward a broader conception of creativity: A case for “mini-c” creativity. *Psychology of Aesthetics, Creativity, and the Arts* 1(2). 73–79.

Benedek, Mathias, Emanuel Jauk, Markus Sommer, Martin Arendasy & Aljoscha C. Neubauer. 2014. Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence* 46. 73–83.

Bergs, Alexander & Nikola A. Kompa. 2020. Creativity within and outside the linguistic system. *Cognitive Semiotics* 13(1). https://doi.org/10.1515/cogsem-2020-2025.

Canovas, Cristobal P. 2020. Learning formulaic creativity: Chunking in verbal art and speech. *Cognitive Semiotics*.

Carroll, John. B. 1993. *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Cattell, James M. 1903. A statistical study of eminent men. *Popular Science Monthly* 62. 359–377.

Diedrich, Jennifer, Emanuel Jauk, Paul J. Silvia, Jeffrey M. Gredlein, Aljoscha C. Neubauer & Mathias Benedek. 2018. Assessment of real-life creativity: The inventory of creative activities and achievements (ICAA). *Psychology of Aesthetics, Creativity, and the Arts* 12(3). 304–316.

Epstein, Robert, Steven M. Schmidt & Regina Warfel. 2008. Measuring and training creativity competencies: validation of a new test. *Creativity Research Journal* 20(1). 7–12.
Feist, Gregory J. 1998. A Meta-Analysis of Personality in Scientific and Artistic Creativity. *Personality and Social Psychology Review* 2(4). 290–309.
Forthmann, Boris, Oluwatosin Oyebade, Adebusola Ojo, Fritz Günther & Heinz Holling. 2018. Application of latent semantic analysis to divergent thinking is biased by elaboration. *The Journal of Creative Behavior* 53(4). 559–575.
Furnham, Adrian & Mark Batey. 2006. Creativity, intelligence, and personality: A critical review of the scattered literature. *Genetic, Social, and General Psychology Monographs* 132(4). 355–429.
Gabora, Liane & Scott Barry Kaufman. 2010. Evolutionary approaches to creativity. In James C Kaufman & Robert Sternberg (eds.), The Cambridge handbook of creativity, 279–300. Cambridge, UK: Cambridge University Press.
Gajda, Aleksandra, Maciej Karwowski & Ronald A. Beghetto. 2017. Creativity and academic achievement: A meta-analysis. *Journal of Educational Psychology* 109(2). 269–299.
Guilford, Joy. P. 1966. Intelligence: 1965 model. *American Psychologist* 21(1). 20–26.
Guilford, Joy P. 1950. Creativity. *American Psychologist* 5. 444–454.
Guilford, Joy P. 1956. The Structure of Intellect. *Psychological Bulletin* (53). 267–293.
Guilford, Joy P. 1960. Basic Conceptual Problems in the Psychology of Thinking. *Annals of the New York Academy of Sciences* 91(1). 6–21.
Hennessey, Beth A. & Teresa M. Amabile. 2010. Creativity. *Annual Review of Psychology* 61(1). 569–598.
Hoffmann, Thomas. 2019. Language and creativity: a Construction Grammar approach to linguistic creativity. *Linguistics Vanguard* 5(1). 1–8.
Hoffmann, Thomas. 2020. Construction Grammar and Creativity: Evolution, psychology and cognitive science. *Cognitive Semiotics* 13(1). https://doi.org/10.1515/cogsem-2020-2018.
Holling, Heinz & Jörg-Tobias Kuhn. 2008. Does intellectual giftedness affect the factor structure of divergent thinking? Evidence from a MG-MACS analysis. *Psychology Science* 50(2). 283–294.
Jäger, A.O., Heinz Martin Süß & A. Beauducel. 1997. *Berliner Intelligenzstruktur-Test: BIS-Test*. Göttingen: Hogrefe.
Jauk, Emanuel, Mathias Benedek, Beate Dunst & Aljoscha C. Neubauer. 2013. The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence* 41(4). 212–221.
Kandler, Christian, Rainer Riemann, Alois Angleitner, Frank M. Spinath, Peter Borkenau & Lars Penke. 2016. The nature of creativity: The roles of genetic factors, personality traits, cognitive abilities, and environmental sources. *Journal of Personality and Social Psychology* 111(2). 230–249.
Karwowski, Maciej, Jan Dul, Jacek Gralewski, Emanuel Jauk, Dorota M. Jankowska, Aleksandra Gajda, Michael H. Chruszczewski & Mathias Benedek. 2016. Is creativity without intelligence possible? A Necessary Condition Analysis. *Intelligence* 57. 105–117.
Kaufman, James C., John Baer, David H. Cropley, Roni Reiter-Palmon & Sarah Sinnett. 2013. Furious activity vs. understanding: How much expertise is needed to evaluate creative work? *Psychology of Aesthetics, Creativity, and the Arts* 7(4). 332–340.

Kaufman, James C. & Ronald A. Beghetto. 2009. Beyond Big and Little: The Four C Model of Creativity. *Review of General Psychology* 13(1). 1–12.
Kelley, Truman L. 1927. *Interpretation of Educational Measurments*. Yonkers NY: World Book.
Kim, Kyung H. 2005. Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education* 16(2-3). 57–66.
Landauer, Thomas K, Peter W. Foltz & Darrell Laham. 1998. An introduction to latent semantic analysis. *Discourse Processes* 25(2–3). 259–284.
Martindale, Colin & Audrey Dailey. 1996. Creativity, primary process cognition and personality. *Personality and Individual Differences* 20(4). 409–414.
McCrae, Robert R. 1987. Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology* 52(6). 1258–1265.
Mumford, Michael D. 2003. Where Have We Been, Where Are We Going? Taking Stock in Creativity Research. *Creativity Research Journal* 15(2-3). 107–120.
Owens, William. A., Charles. F. Schumacher & J. B. Clark. 1957. The measurement of creativity in machine design. *Journal of Applied Psychology* 41(5). 297–302.
Parkhurst, Howard B. 1999. Confusion, Lack of Consensus, and the Definition of Creativity as a Construct. *The Journal of Creative Behavior* 33(1). 1–21.
Patrick, Catharine. 1935. Creative thought in poets. *Archives of Psychology* 26. 1–74.
Plucker, Jonathan A., Ronald A. Beghetto & Gayle T. Dow. 2004. Why Isn't Creativity More Important to Educational Psychologists? Potentials, Pitfalls, and Future Directions in Creativity Research. *Educational Psychologist* 39(2). 83–96.
Polya, George. 1945. *How to solve it: a new aspect of mathematical method* (Princeton Science Library). Expanded Princeton Science Library ed. Princeton [N.J.]: Princeton University Press.
Prabhakaran, Ranjani, Adam E. Green & Jeremy R. Gray. 2014. Thin slices of creativity: Using single-word utterances to assess creative cognition. *Behavior Research Methods* 46(3). 641–659.
Preckel, Franzis, Christina Wermer & Frank M. Spinath. 2011. The interrelationship between speeded and unspeeded divergent thinking and reasoning, and the role of mental speed. *Intelligence* 39(5). 378–388.
Puryear, Jeb S., Todd Kettler & Anne N. Rinn. 2017. Relating Personality and Creativity: Considering What and How We Measure. *The Journal of Creative Behavior* 53(2). 232–245.
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.
Reiter-Palmon, Roni, Boris Forthmann & Baptiste Barbot. 2019. Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts* 13(2). 144–152.
Renzulli, Joseph S, Linda H Smith, Alan J White, Carolyn M Callahan, Robert K Hartman & Karen L Westberg. 2002. *Scales for rating the behavioral characteristics of superior students. Technical and administration manual*. Creative Learning Press.
Rhodes, Mel. 1961. An Analysis of Creativity. *The Phi Delta Kappan* 42. 305–310.
Richards, Ruth, Dennis K. Kinney, Maria Benet & Ann P. Merzel. 1988. Assessing everyday creativity: Characteristics of the Lifetime Creativity Scales and validation with three large samples. *Journal of Personality and Social Psychology* 54(3). 476.
Root-Bernstein, Robert S. 1984. Creative Process as a Unifying Theme of Human Cultures. *Daedalus* 113. 197–219.

Runco, Mark A. 2008. Commentary: Divergent thinking is not synonymous with creativity. *Psychology of Aesthetics, Creativity, and the Arts* 2(2). 93–96.
Runco, Mark A. & Garrett J. Jaeger. 2012. The Standard Definition of Creativity. *Creativity Research Journal* 24(1). 92–96.
Sampson, Geoffrey. 2006. Two ideas of creativity. *Evidence, Experiment and Argument in Linguistics and Philosophy of Language* 11.
Silvia, Paul J., Roger E. Beaty & Emily C. Nusbaum. 2013. Verbal fluency and creativity: General and specific contributions of broad retrieval ability (Gr) factors to divergent thinking. *Intelligence* 41(5). 328–340.
Silvia, Paul J., James C. Kaufman & Jean E. Pretz. 2009. Is creativity domain-specific? Latent class models of creative accomplishments and creative self-descriptions. *Psychology of Aesthetics, Creativity, and the Arts* 3. 139–148. doi:https://doi.org/10.1037/a0014940.
Silvia, Paul J., James C. Kaufman, Roni Reiter-Palmon & Benjamin Wigert. 2011. Cantankerous creativity: Honesty–Humility, Agreeableness, and the HEXACO structure of creative achievement. *Personality and Individual Differences* 51(5). 687–689.
Silvia, Paul J., Beate P. Winterstein, John T. Willse, Christopher M. Barona, Joshua T. Cram, Karl I. Hess, Jenna L. Martinez & Crystal A. Richard. 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(2). 68–85.
Simonton, Dean K. 2012. Taking the U.S. Patent Office Criteria Seriously: A Quantitative Three-Criterion Creativity Definition and Its Implications. *Creativity Research Journal* 24(2-3). 97–106.
Torrance, Ellis. P. 1966. *The Torrance Tests of Creative Thinking—Norms, Technical Manual Research Edition—Verbal Tests, Forms A and B—Figural Tests, Forms A and B*. Princeton [N.J.]: Personnel Press.
Trousdale, Graeme. 2020. Creativity, reuse, and regp:ularity in music and language. *Cognitive Semiotics* 13(1). https://doi.org/10.1515/cogsem-2020-2021.
Turner, Mark. 2020. Constructions and Creativity. *Cognitive Semiotics*.
Uhrig, Peter. 2020. Creative Intentions – The fine line between 'creative' and 'wrong.' *Cognitive Semiotics* 13(1). https://doi.org/10.1515/cogsem-2020-2027.
Vincent-Lancrin, Stéphan. 2017. *Future learning skills in a global context: from discourse to practice*. http://docplayer.net/51475061-Future-learning-skills-in-a-global-context-from-discourse-to-practice.html.
Wallach, Michael A. & Nathan Kogan. 1965. A new look at the creativity-intelligence distinction1. *Journal of Personality* 33(3). 348–369.
Weiss, Selina, OliverWilhelm & PatrickKyllonen. submitted. *A Review and Taxonomy of Creativity Measures*.
Wertheimer, M. 1945. *Productive Creativity*. New York: Harper & Row.
Wilhelm, Oliver, Andrea Hildebrandt & Klaus Oberauer. 2013. What is working memory capacity, and how can we measure it? *Frontiers in Psychology* 4.
Wilson, Robert C., Joy P. Guilford & Paul R. Christensen. 1953. The measurement of individual differences in originality. *Psychological Bulletin* 50(5). 362–370.

Bionotes

Selina Weiss
Ulm University, Ulm, Germany
selina.weiss@uni-ulm.de
Selina Weiss is a doctoral student of psychology at Ulm University, Germany. Her dissertation focuses on individual differences in creativity, including competing theoretical accounts and available psychometric measures. She has worked with several renowned scientists from this field and has conducted several large multivariate studies. In addition to her research on creativity, her research interests lie in novel assessment approaches for creativity, comprehension, and interpersonal skills and abilities.

Oliver Wilhelm
Ulm University, Ulm, Germany
oliver.wilhelm@uni-ulm.de
Oliver Wilhelm is Professor of Psychology at Ulm University, Germany. He earned his doctoral degree in 2000 and worked as postdoctoral fellow at the Georgia Institute of Technology and the University of Arizona. He subsequently worked as assistant and associate professor at Humboldt University Berlin and as full professor at the University of Duisburg-Essen. He studies individual differences in human abilities such as working memory, reasoning, emotion perception, and scholastic achievement.

Special Issue: *Construction Grammar and Creativity* edited by Thomas Hoffmann

III. Manuscript II

A Review and Taxonomy of Creativity Measures

The idea and design were developed by Prof. Dr. Oliver Wilhelm, Dr. Patrick Kyllonen, and myself. The literature review and discussion of results were conducted by myself and supervised by Prof. Dr. Oliver Wilhelm and Dr. Patrick Kyllonen. The rating study and its analysis were performed by myself. I wrote the manuscript and the revisions, and all co-authors revised these versions. In its current form, the manuscript is submitted for possible publication in the journal Psychology of Aesthetics, Creativity, and the Arts on the 8th of December 2020. Draft version ACA-2020-0580, 03/18/20. This paper has not been peer reviewed. Please do not copy or cite without author's permission.

Weiss, S., Wilhelm, O., & Kyllonen, P. (under review, submitted 2020). A review and taxonomy of creativity measures. *Psychology of Aesthetics, Creativity, and the Arts*.

A Review and Taxonomy of Creativity Measures

Selina Weiss[1*], Oliver Wilhelm[1], & Patrick Kyllonen[2]

[1] Department of Individual Differences and Psychological Assessment, Institute for Psychology and Education, Ulm University, Ulm, Baden-Wuerttemberg, Germany

[2] Educational Testing Service, ETS, Princeton, New Jersey, United States of America

* Corresponding author

E-mail: selina.weiss@uni-ulm.de

Author Note:

Correspondence concerning this article should be addressed to Selina Weiss, Institute for Psychology and Pedagogy, Ulm University, Albert-Einstein Allee 47, 89081 Ulm, Germany. Telephone: +49-(0)731/50 31145. Fax: +49-(0)731/50 31149. Email: selina.weiss@uni-ulm.de

This research was supported was not supported by any grants. We thank the research assistants who helped coding the measures. For all measures, we have reported example items, besides we report data exclusion criteria. The sample items along with the human ratings are available in the OSF https://osf.io/e49n3/?view_only=2fb7338339464fb3acf2dc6a3e26b612.

Abstract

There is growing interest in creativity as evidenced by employers' surveys of skills sought, future workforce skills requirements, and an increased emphasis on education. This indicates a need for applied creativity assessments (such as college admissions and workforce selection) and research on the psychometric soundness of creativity indicators. However, the assessment of creative abilities presents major challenges. The many competing and complementary ideas on measuring creativity have resulted in a wide diversity of measures, making it difficult for potential users to decide on their appropriateness. Prior research has proposed creativity assessment taxonomies, but we argue that these have shortcomings because (a) they have not been based on the essential assessment features, and (b) they are insufficiently specified for reliably categorizing extant measures. We propose a new framework for categorizing creativity measures based on the following attributes: (a) measurement approach (self-report, other-report, ability tests), (b) construct (e.g., creative interests and attitudes, creative achievements, divergent thinking), (c) data type generated (e.g., questionnaire data vs. accomplishments counts), (d) prototypical scoring method (e.g., consensual assessment technique), and (e) psychometric problems. We identified 213 creativity measures appearing in the literature since 1900 and classified each measure according to their task attributes by two independent raters (rater agreement Cohen's kappa .83 to 1.00 for construct). We provide a summary of convergent validity evidence and psychometric shortcomings. We conclude with recommendations for using the taxonomy and some psychometric desiderata for future research.

Keywords: Creativity; Measurement; Review; Taxonomy

Most people believe they know what creativity is and how it manifests. Some of these ideas are idiosyncratic. What seems to be creative to one person may not seem creative to another and might differ from what is identified as creative in the scientific literature (Dawson et al., 1999). Nevertheless, creativity is present in everyday life (e.g., a home chef inventing new recipes while cooking), in schools (e.g., a student creating ideas for more sustainable energy use), and in work-life (e.g., an engineer inventing a non-breakable smartphone). Such creative ideas and products—including ideas from diverse domains such as medicine, technology, and philosophy—might have provided the human species an evolutionary advantage over other species (Gabora & Kaufman, 2010).

The importance of creativity in work life is stressed by a recent survey among world business leaders from 79 countries (PWC, 2016, January). This survey describes creativity as an essential supplement in a machine-driven and computerized world as computers will automate many jobs. At the same time, the creative achievement is perceived as currently being outside the realm of computers. This makes creativity highly valuable as a future workforce skill and a key attribute in recruiting new employees (Gray & Koncz, 2017, November). Next, this underscores the importance of preparing students for these future work requirements and challenges with creative skills, such as creative problem-solving and thinking (Bughin et al., 2018, May 23; Craft, 2005; NACCCE, 1999). Hence, the educational sector focuses more and more on understanding (e.g., Lucas et al., 2013), measuring (Vincent-Lancrin, 2017, May), and fostering creativity (e.g., Berg et al., 2012). Although it is important, teaching creativity in schools is challenging, and schools often focus on traditional curricular domains rather than on creativity per se (Westby & Dawson, 1995; Kaufman & Plucker, 2011).

Despite a long history of creativity research and the great interest in understanding and measuring creativity, the question of how best to assess creativity in a psychometrically sound manner is unresolved (Barbot et al., 2019; Long et al., 2014; Plucker & Makel, 2010; Parkhurst, 1999; Plucker et al., 2004; Runco, 2004). The present review summarizes and describes measurement attempts and provides measurement attributes for an exhaustive categorization system – a measurement taxonomy. This taxonomy supports the choice of specific measures for basic or applied research and various practical purposes.

Uses of Creativity Assessment

There are two reasons for focusing on creativity assessment. One is that to understand creativity it is necessary to measure it. For example, experiments designed to boost creativity can be evaluated with pre-post measures of creativity. In this case, formative assessments or feedback can be used to help students develop and improve their creative skills based on individualized instruction (Lucas et al., 2013). A means to better understand the importance of creativity is the evaluation of intervention programs (e.g., Alfonso-Benlliure et al., 2013) and large-scale comparative assessments (Barbot, et al., 2019). Large-scale comparative assessments that include measures of creativity (e.g., creative thinking), such as PISA 2021 (OECD: PISA 2021; Vincent-Lancrin, 2017, May) allow for a comparison of different schools, districts, and economies and enable the monitoring of creativity growth over time, which leads to a better understanding of creativity (Kim, 2011).

A second reason to be concerned with the measurement of creativity is that creativity assessments have practical utility. They can be used in high-stakes contexts, often as summative assessments, such as higher education admissions, workforce selection, skills certification, and school comparisons. For example, the Rainbow Project and the Kaleidoscope Project (Sternberg, 2009) focused on creative idea generation as potential admissions assessment measures for undergraduate admissions.

Challenges of Creativity Assessment

Creativity assessment comes with unique challenges. A key challenge is defining creativity. Although there is widespread consensus regarding the definition of creativity, this definition is quite broad. Experts mostly agree on a bipartite definition that includes an original product (new, unusual, novel, unexpected) that is somehow valuable (useful, good, adaptive, appropriate) (e.g., Barron, 1955; Batey, 2012; Mumford, 2003; Runco & Jaeger, 2012; Stein, 1953).

Even though originality and value are widely accepted, they might not capture every aspect of creativity, and both are hard to define, evaluate, and measure (Abraham, 2018). For example, the value of highly creative products (e.g., in the arts) may be subjective and based on societal norms (Runco & Jaeger, 2012, Plucker et al., 2004). Originality can have different meanings in different contexts (e.g., original to whom?). Big-C creativity means original to the world (Simonton, 1977, 1998), whereas small-c creativity is a lower bar referring to original to the examinee (Richards et al., 1988). Depending on the application, either might be appropriate—judging art and music creativity in world-class artists calls for the use of Big-C creativity; judging creativity of a high school student's art project more appropriately relies on small-c as an evaluative reference.

Because of this vague definition, the boundaes between creativity and other constructs, such as problem-solving (Weisberg, 1988), openness to experience (Feist, 1998; McCrae, 1987), and general retrieval ability (Forthmann, et al., 2019) present the third challenge. Previous research suggests a modest relationship between creativity and IQ-scores (r = .17; 95% CI [.17, .18]; N = 45,880; Kim, 2005) and a small correlation between creativity and academic achievement (r = .22; 95% CI [.19, .24]; N = 52,578; Gajda et al., 2017). Similarly, the literature indicates that creativity is related to the personality factor of openness to experience, with a medium-sized strength (divergent thinking tests and self-reported Openness to experience, r = .39; McCrae, 1987). Hence, creativity seems to share its "no man's land" status (Stankov, 1999) between intelligence and personality with other constructs such as emotional intelligence.

Previous Categorization of Creativity Assessment

The creativity literature provides several taxonomies for categorizing research and measurement (e.g., Snyder et al., 2019). Extant taxonomies vary widely concerning terminology, as they are often based on different theoretical approaches and serve different purposes. Therefore, we first provide an overview applying the original terminology describing the most prominent taxonomies. One categorization approach—which has been further applied and developed in later taxonomies—describes the process of creativity, the creative product, the creative person, and the environment in which creativity arises (4 P's; Rhodes, 1961). This approach to understanding creativity has been further applied and developed as a categorization basis (e.g., Batey, 2012; Plucker & Makel, 2010; Snyder et al., 2019; Said-Metwaly et al., 2017). A systematic review by Snyder and colleagues (2019) included measures from 1984 to 2013 that were coded by the type of measurement (e.g., self-rating of creativity). Next, the measurement type was related to one or more categories of the 4 P's. Another approach, including the 4 P's as facets, describes the measurement approach (e.g., self-rating) and the observational unit where creativity is measured (e.g., team).

Other taxometric approaches categorized measurements either very broadly by the research purpose (e.g., case-studies: Long, 2014; Long & Plucker, 2015; biological: Mayer, 1999) or describe measurement categories. The described measurement categories vary in their breadth and overlap. Hocevar (1981) described ten measurement categories (such as divergent thinking, product judgment), while Lubart (1994) described eight such categories (including peer-reports, divergent thinking) and Barbot (2019) classify indicators as divergent thinking, product assessment, and self-reports. Divergent thinking tasks are particularly prominent in the literature and appear in many taxonomies—either described as measurement approach, construct, or even data type (e.g., Batey, 2012; Barbot, et al., 2019; Plucker & Makel, 2010). Besides the measurement type, other research focusses on the measurement content describing, for example, different dimensions of creative activities (e.g., Kirschenbaum, 1998), different constructs (imagination, creativity, and innovation; Forgeard & Kaufman, 2016), or different foci (e.g., activities; Kaufman, 2019). Other prominent approaches describe the assessment technique (e.g., product assessment; Amabile, 1996), data distributions (Eysenck, 1996), and nomological net (Mumford & Gustafson, 1988). Recent research also focuses on various interdisciplinary approaches and broadens our understanding by including neuroscience methodologies for studying creativity (Benedek et al., 2019).

Limitations of Prior Categorizations

Even though various taxometric approaches exist, they are difficult to align because they are based on different terminology or only address creativity assessment from a specific angle. The jingle-jangle fallacy (Kelley, 1927, p. 64)—that different constructs carry the same name or a common construct has different labels—is a relevant issue in creativity research. For example, the content measured in creativity tests is variously described as constructs (Forgeard & Kaufman, 2016), focus (Kaufman, 2019), creative-activity (Kirschenbaum, 1998), or facet (Batey, 2012). The facet described in Batey (2012) further refers to Rhodes 4 P's (1961), but instead of the dimension, person, he introduces the dimension, trait.

A second limitation of many prior taxonomy attempts is that they describe creativity from a specific focus (e.g., product assessment; Amabile, 1996; self-report; Forgeard & Kaufman, 2016). This particular focus helps specific applications (e.g., neuroscience; Benedek, et al., 2019) but is not sufficient for an overall categorization system designed to comprehensively classify creativity assessments. Here we argue that a comprehensive taxonomy with a common terminology is useful, mainly when it is broad enough to include varying strands of research and unambiguous enough to allow for distinct categorizations.

Purpose of our Review

The purpose of this study is to assemble a broad, comprehensive list of creativity measures that have been used in the literature and published since 1900. We organize this list by specifying a set of features based on our review and extension of previous taxonomies, enabling a comprehensive and unambiguous categorization of measures. We aim to answer the following questions:

1. Which task attributes can be derived from previous creativity taxonomies?
2. Which creativity measures have been used to assess creativity in contexts such as admission decisions in an educational context, personal selection in an I&O context, and large-scale assessment since 1900?
3. Can these measures be categorized reliably according to the specified task attributes?
4. What are the general limitations and psychometric problems of measurement approaches or proposed creativity constructs? How can these problems be solved? Which recommendations can be derived for future research?

Method

Search Strategy

As the study aimed to build a taxonomy that allows a reliable and exhaustive categorization, we have conducted a focused literature review. This includes an in-depth search that followed a predefined protocol to identify a larger number of creativity measures. We searched the creativity literature for creativity measures published in English across a large number of domains (e.g., verbal creativity, arts, music) and groups (e.g., gifted children, adults). We searched online using Google Scholar, Web of Science, Research Gate, and PsycINFO. We also searched the Educational Testing Service (ETS) Test collection, a fairly comprehensive test database that lists over 25,000 tests and measurement devices from the early 1900s to the present. We identified tests based on the following keywords that were used keyword by keyword, similarly in all search engine: *measures, assessment, questionnaire, test, creative, creativity, divergent thinking, flexibility, fluency, originality, ingenuity, gifted child, talent, eminent people, biographical, creative achievements, creative behavior, -interests, -attitudes, - styles, -personality.* Additionally, we searched in previous reviews, special issues, and book chapters on creativity and its measurement, applying forward and backward search through the cited and citing literature. This expands inclusion to measures not represented by the keywords.

We included measures ranging from single-item to numerous items and scales, and we did not screen for test quality. We excluded tests and articles that were limited in at least one of the following ways: (a) was published only in languages other than English, (b) assessed constructs only related to creativity—such as innovation (a key term that is primarily prevalent in business journals; Forgeard & Kaufman, 2016)—or problem-solving. Although these constructs are theoretically related to creativity, we assume that they should be understood as covariates and not as indicators of creativity; (c) were not published or at least applied in a (larger) empirical study that would allow an assessment of psychometric properties; and (d) did not provide item examples, which are obviously indispensable for further evaluation in this review.

This search resulted in a set of 213 different creativity tests. We acknowledge that due to the exclusion criteria, tests that some researchers might regard as creativity tests are not included.

Results

Task Attributes Derived from Previous Categorizations

We aim to provide a taxonomy that considers all hitherto discussed task attributes and allows an unambiguous categorization of measurement. Table 1 summarizes prominent taxonomies and categorization systems presenting the original labels. Based on these labels, we derived a unifying terminology. The task attributes we identified from previous taxonomies allow a:

a) broad and general categorization of creativity assessments by describing the general measurement approach

b) a specific description of the content that is typically measured by defining a construct

c) description of the data type of each measurement approach

d) recommendation for scoring each measurement type

e) summary of psychometric problems associated with the attributes described above.

Table 1

Previous Taxonomies with Topic Coverage

	Original categorization/taxometric approach	Primarily included attributes
Rhodes, 1961	process, product, person, press	⏱ 🗪
Hocevar, 1981	ten measurment categories (e.g., divergent thinking, product judgments)	⏱ 🗪 α
Mumford & Gustafson, 1988	nomological net of creative potential (e.g., environment, skills/abilities, and personality)	🏆
Lubart, 1994	eight measurement categories (e.g., peer reviews)	⏱ 🗪
Eysenck, 1996	trait (e.g., personality and ability) or achievement measures	🗪 🏆
Amabile, 1996	assessment techniques (e.g., tests, subjective judgment)	≜
Kirschenbaum, 1998	nine dimenisions of creative activity (e.g., interest)	🗪
Mayer, 1999	psychometric, psychological, biographical, biological, computational, contextual	⏱
Plucker & Makel, 2010	process, person, enviroment	⏱ 🗪 ≜ α
Batey, 2012	facets (e.g., process, product), measurement approach (e.g., self-rating), and level (e.g., individuals)	⏱ 🗪 🏆
Long, 2014; Long & Plucker, 2015	case studies, experimental paradigms, and questionnaire approaches	⏱ 🗪 ≜
Forgeard & Kaufman, 2016	constructs (e.g., imagination, creativity, and innovation)	🗪 α
Barbot, et al., 2019	Divergent thinking, product assessment, self-reports	⏱ 🗪 ≜ α
Benedek et al., 2019	task-type (e.g., problem-solving) and neuroscience method	⏱ 🗪 ≜ α
Kaufman, 2019	focus (e.g, activties, evaluation, process)	🗪 α
Snyder et al., 2019	measurement types (e.g., self-, and other measures) and category (e.g., process, person)	⏱ 🗪 ≜
Said-Metwaly, et al., 2017	process, product, person, press	⏱ 🗪 α

Note. ⏱ = measurement approach; 🗪 = construct; 🏆 = data type; ≜ = scoring; α = psychometric problems.

Table 1 includes the previous taxonomies' original terminology and shows which of the aforementioned attributes are already represented in the taxonomy. As an example: The measurement types (Snyder, et al., 2019), task types (Benedek et al., 2019), measurement approaches (e.g., Batey, 2012), and measurement categories (e.g., Hocevar, 1981) all refer to the same attribute: the measurement approach (e.g., self-report). The attribute "construct" describes the underlying traits (e.g., imagination, Forgeard & Kaufman, 2016; beliefs, Kaufman, 2019) or abilities measured (e.g., divergent thinking). Eysenck's taxonomy (1996) includes a particular categorization by data type, as he points out that scores on divergent thinking tasks tend to be normally distributed. In contrast, measures of creative achievements, such as criterion measures, which are typically count variables, are better captured by a highly skewed Poisson distribution as few people receive such recognitions (Eysenck, 1996; Lotka, 1926). As the measurement of creativity is complex, a comprehensive taxonomy should discuss scoring approaches and the psychometric problems of the reviewed tests. Some reviews have stressed general psychometric problems such as predictive, discriminant, and construct validity (e.g., Benedek, et al., 2019; Plucker & Makel, 2010); this, for example, resulted in recommending open-science item pools for creativity assessment, Barbot, et al., 2019).

A Proposed Taxonomy of Creativity Assessment

Based on previous frameworks (see Table 1), we propose a taxonomy, including theoretically derived task attributes that combine and complete previous categorization approaches. Table 2 presents our taxonomy.

Table 2

A Taxonomy of Measuring Creativity, including Task Attributes.

	Self-report				Other-report				Ability Tests	
	Preferences		Accomplishments		Preferences		Accomplishments		Performance	
Construct	Creative interests and attitudes	Creative Personality (e.g., Ingenuity)	Creative achievements (e.g., on the job)	Creative activities (e.g., cooking)	Creative attitudes & characteristics	Diver-gent think-ing	Archival ratings (e.g., eminent people)	Original outcomes (e.g., patent indices)	Divergent thinking (fluency, flexibility, originality)	Insight problems
Data type	Questionnaire-data (Cattell, 1958): Subjective data		Life-data (Cattell, 1958): Biographical data		Questionnaire-data (Cattell, 1958): Subjective data		Life-data (Cattell, 1958): Objective data		Test-data (Cattell, 1958): Objective data	
Prototypical Scoring	Average scale scores		Sum of self-nominations (rating-scale of accomplishments)		Averaging scale scores/Agreement of raters		Agreement of raters/experts		Divergent thinking: Fluency, flexibility, originality assessed by agreement of raters	Insight: proportion attainment of objectively correct responses
Psychometric Problems	Introspection of the individual; Highly fakeable; Shortcomings of self-report; Evaluation of originality, novelty, and fluency; Convergent Validity		Introspection of the individual; Fakeable; Shortcomings of self-report; Unusual data distribution; Convergent validity		Convergent validity; Inter-rater reliability: familiarity with creativity and familiarity with the target; Artifact of the assessment object		Inter-rater reliability; Bipartite definition of creativity: usefulness and novelty; Domain-specific		Divergent Thinking: Inter-rater reliability; intelligence and speed as sources of variance	Insight: Assessmen t of novelty and originality

It presents three major measurement approaches: self-report measures, other-report measures, and ability tests. These measurement approaches, at the apex of the hierarchy, offer a broad disjunct categorization. The measurement approaches are divided by preferences and accomplishments (self-, and other-reports) and performance to narrow this categorization. These subcategories are derived from different task requirements (preferences: report your/others typical behavior; accomplishments: frequency/rating of (previous) performances; Performance: produce creative ideas/solve a creative problem). Each of these measurement approaches subcategories is to come with further task attributes—constructs, data types, prototypical scoring, and specific psychometric problems—allowing a more precise categorization of a measure. Categorization of an assessment by measurement approach and construct allows an easy and straightforward categorization. The other task attributes offer a descriptive tool nested below the measurement approaches. The different data types usually come with a specific behavior of the participant and response style (Questionnaire data: Typical behavior on a Likert Scale; Life-data: Maximal effort/skill-related typical behavior on a checklist; and Test-data: Maximal cognitive effort in open-ended answers). This leads to prototypical scoring recommendations and specific limitations. Before we further describe the constructs and other attributes of the taxonomy, we present an overview of the identified measures.

Measures Identified

Figure 1 shows an overview of the *N* = 213 measures identified in the literature review and their categorization by measurement approach and construct. Please note that a comprehensive list of all measures, including sample items and the original source of the measures, is available in the OSF (https://osf.io/e49n3/?view_only=2fb7338339464fb3acf2dc6a3e26b612). Ability tests are most prominent in the literature (*n* = 119), followed by self-report measures (*n* = 58) and other report measures (*n* = 36).

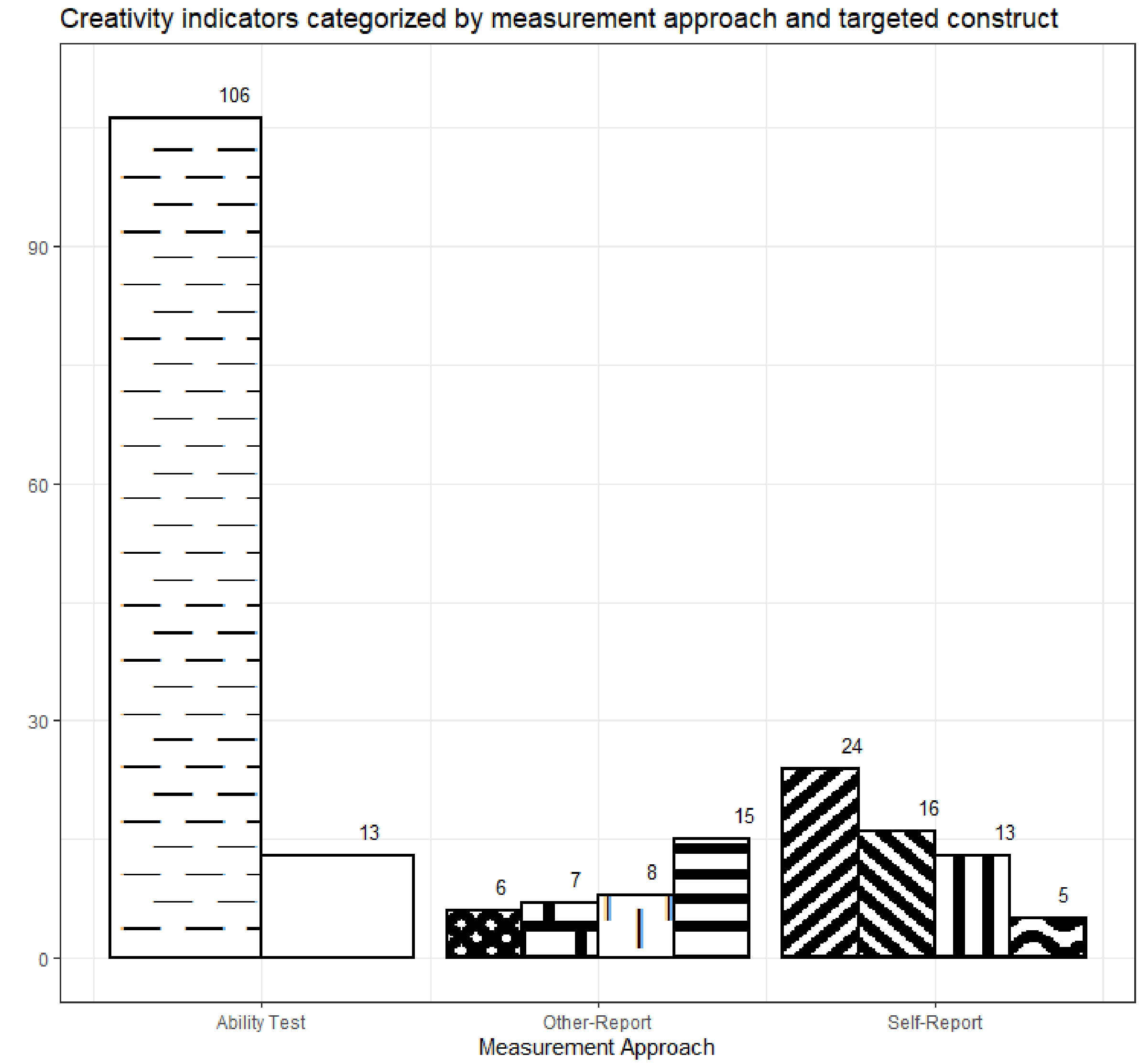


Figure 1. Creativity indicators: = Divergent Thinking, blank = insight; = Attitudes & characteristics, = Divergent Thinking, = Archival, = Outcomes, = Interest and attitudes, = Personality, = Achievements, and = Activities.

Rater Study: Identified Measures

Two independent human raters categorized all tests presented in Figure 1 with respect to their construct. The raters were provided with a general description of all constructs and a table including the test name, the author, and a minimum of one sample item (e.g., Write a paragraph about the theme “a tree”; Ekstrom, et al., 1976). Based on this table, the raters rated

(0 = construct does not apply; 1 = construct does apply). These results are provided in the OSF (https://osf.io/e49n3/?view_only=2fb7338339464fb3acf2dc6a3e26b612). After the rating, the agreement between the raters was analyzed based on Cohen´s kappa (Cohen, 1960). After the rating, the raters' agreement was analyzed based on Cohen's kappa (Cohen, 1960). Table 3 displays the kappa for categorizing the construct, along with the percentage of agreement. In a further meeting of raters' consensus for all categories was reached.

Table 3

Cohen´s Kappa (κ) and Percentage of Agreement for Two Independent Raters for the Measurement Construct and Construct.

Construct	Cohen´s Kappa (κ)	Percentage of Agreement
Self-report (Construct: Interest & Attitudes)	.84	97.2%
Self-report (Construct: Personality)	.83	97.2%
Self-report (Construct: Achievements)	.92	99.1%
Self-report (Construct: Activities)	.83	99.1%
Other-report (Construct: Divergent Thinking)	1.00	100%
Other-report (Construct: Attitudes & Interests)	1.00	100%
Other-report (Construct: Archival)	1.00	100%
Other-report (Construct: Outcome)	1.00	100%
Ability Test (Construct: Divergent Thinking)	.99	99.5%
Ability Test (Construct: Insight)	1.00	100%

Categorizing Measures based on the Proposed Taxonomy

In the following paragraphs, we provide a detailed description of the attributes that are presented in Table 2. Further, we describe the identified measures, including sample items for each measure. We also discuss psychometric shortcomings for the measurement approaches.

Constructs in Self-Reports

First, measuring creative interests and attitudes can be seen as a self-report construct (e.g., Hocevar, 1981). Creative attitudes can be defined as the degree of positive affect regarding creative dispositions, temperaments, and orientation, and predispose one to think creatively about a product (Davis, 1999). A person who shows creative interests enjoys being involved in creative activities (Slahova et al., 2007). These constructs are not further distinguished from one another as they lack clear segregation based on empirical data.

Second, the study of personality investigates individual differences in behavior and thinking (Kazdin, 2000). Therefore, it is useful to consider the evidence for whether a creative personality can be found within personality assessments. The literature often relates the personality factor openness to experience with creativity. Some openness measures, such as the ingenuity scale, include self-reported creativity (Woo et al., 2014).

Third, the construct of everyday creativity in self-reports is also known as small-c (Richards et al., 1988) describing minor creative accomplishments in life that can happen every day and enhance and enrich our everyday lives.

Fourth, previous creative achievements, a different construct assessing accomplishments, can be applied in various areas such as language, arts, science, and social studies (Torrance, 1962; Lees-Haley, 1978; Bull & Davis, 1980).

Data Type and Scoring in Self-Reports

Self-reports are either based on the data type Q-data (Cattell, 1958) and hence on subjective data or L-data (Cattell, 1958), including biographical data. Q-data approaches are probably the most common form of assessment in psychology due to the ease of item development, scoring, and analysis of Likert-style items. L-data are gathered to describe previous creativity and creative behavior in various contexts based on the participants' introspection and self-evaluation. They can either include multiple-choice items, Likert-scale items (e.g., ranging from "never" to "more than ten times"; Diedrich et al., 2018) or open-ended answers (e.g., "Please name your overall five most creative achievements in your life so far.", Diedrich et al., 2018). The participants are asked to report the frequency of previous creative actions and performances or provide information about the production of creative things. Q-data requires the participants' typical behavior, whereas L-data indicates skill-related typical behavior or the report of maximal cognitive effort. Skill-related typical behavior includes behavior that requires a high degree of expertise (e.g., playing concert piano), but if this degree of expertise is reached, the behavior is more akin to typical behavior than maximal cognitive effort.

Self-Report Measures

In all Q-data measures in self-reports, the participants were asked to answer based on their typical behavior. Table 4 presents questionnaires, including example items that assess creative interests and attitudes. A measure used with elementary-school-age children and therefore of interest for teachers is the Creative Attitude Survey, which includes scales about imagination, interest in art and writings, attraction to abstract and magical themes, and preference for novelty (Schaefer & Bridges, 1970), the Pennsylvania Assessment of Creative Tendency (Rookey, 1971) and efficacy scale (e.g., creative productivity to identify gifted children; Schack, 1989).

Table 4

Self-report (Preferences): Measures of Creative Interests and Creative Attitudes.

Measure	Author	Example Items
Study of Values	Allport et al., 1960	If you had some time to spend in a waiting room and there were only two magazines to choose from, would you prefer: (a) Scientific Age; (b) Arts and Decorations?
Creative Interests subscale (Guilford-Zimmermann Interest Inventory)	Guilford, 1963	I often write words in new combinations to convey emotion rather than meaning.
Opinion, Attitude, and Interest Survey	Fricke, 1965	Rules and regulations often do more harm than good.
Rating scale of Creative Characteristics	Meeker, 1978; Guilford, 1967	Stimulated by reading
Creative Attitude Survey	Schaefer & Bridges, 1970	I get bored easily.

Pennsylvania Assessment of Creative Tendency	Rookey, 1971	Creating new words is dumb.
Creative Behavior Disposition Scale	Taylor et al., 1974; Taylor & Fish, 1979	I combine things in a new way.
How Creative are You? Inventory	Raudsepp, 1981	Daydreaming has provided the impetus for many of my more important projects.
Group Inventories for finding creative talent/ Group Inventories for fining Interest	Rimm, 1980, Davis & Rimm 1982	I like to make up my own songs.
Efficacy scale for creative productivity	Schack, 1989	I usually work on projects that are not a class assignment if it is about something I am interested in.
Creative Styles Questionnaire	Kumar & Holman, 1989	I typically create new ideas by systematically modifying (by substituting, rearranging, elaborating, etc.) an existing idea.
Khatena-Morse Multitalented Perception Inventory	Khatena & Morse, 1994	I have a fluent and vivid imagination which I use in accomplishing tasks.
Abedi–Schumacher Creativity Test	O´Neil et al., 1994; Auzmendi et al., 1996	Approaching a complex task, I come up with a variety of approaches.
Runco Ideational Behavior Scale	Runco et al., 2001	Quality of ideas is more important than quantity.
Creative Self-Efficacy scale	Tierney & Farmer, 2002	I have confidence in my ability to solve problems creatively.
Epstein Creativity Competencies Inventory for Individual	Epstein et al., 2008	I always keep a recording device by my bed at night. (Capturing) When I set goals for myself, I make sure they are open-ended. (Challenging) I often read books from outside my specialty. (Broadening) I redecorate or rearrange my work environment regularly. (Surrounding)
Cognitive Processes Associated with Creativity	Miller, 2009	While working on a problem, I try to generate as many ideas as possible.
Creative Process Engagement Scale	Zhang & Bartol, 2010	I consider diverse sources of information in generating new ideas.
Kaufman Domains of Creativity Scale	Kaufman, 2012	Finding something fun to do when I have no money.
Creative Cognition Scale in studying	Rogaten & Moneta, 2015	Incorporating previous solutions in new ways leads to good ideas.
Team Creativity Climate	Kiratli et al., 2016	In our team we are open to each other´s views and ideas.
Reisman Diagnostic Creativity Assessment	Reisman et al., 2016	I can generate many solutions.

The Mode Shifting Index	Pringle & Snowden, 2017	When working on a task, I like to think both in depth about the details and drift out of focus and let my mind wander (e.g. looking out of the window).

Note. Tests are ordered by publication year. The participant should either rate their agreement on a Likert scale or just agree or disagree. All example items display items either loading high on a factor/ subscale or are theoretically relevant for the measurement construct. A few items were rephrased.

Table 5 presents measures that were explicitly designed to measure creative personality. These measures are mostly based on adjective checklists (Domino, 1994) or sentences that describe artistic areas (e.g., areas of artistic inclination, intelligence, imagination, self-confidence; Khatena-Torrance Creative Perception Inventory; Khatena & Torrance, 1976). Besides, many studies are based on ingenuity items or other subscales from personality inventories. The International Personality Item Pool includes items on creativity and ingenuity (Goldberg et al., 2006), as well as the Abridged Big 5 Circumplex (e.g., "I am full of ideas."; Hofstee et al., 1992; Johnson, 1994), the Hogan Personality Inventory (e.g., curiosity: "I have taken things apart just to see how they work."; Hogan & Hogan, 1992), the Jackson Personality Inventory-Revised (Jackson, 1994) and the Self-Directed Search questionnaire (Holland et al., 1997).

Table 5

Self-report (Preferences): Measures of Creative Personality

Measure	Author	Example Items
Khatena-Torrance Creative Perception Inventory	Khatena & Torrance, 1976	Using the strategy of restructuring.
Creative Personality Scale	Gough, 1979	Unconventional
Adjective Checklist	Gough & Heilbrun, 1983	Zany
The Creatix Inventory	Byrd, 1986	I am an innovator.
Creative Temperament Scale	Gough, 1992	Inquisitive
Domino Creative Scale	Domino, 1994	Imaginative
Emotional Creativity Inventory	Averill, 1999	I have emotional experiences that would be considered unusual or out of the ordinary.
Creative Personality Scale	Kaufman & Baer, 2004	Love to think up new ways of doing things.
Creative Approach Questionnaire	Durmysheva & Kozbelt, 2010	I begin projects without a detailed understanding of where it will lead me.
Thinking and Creative Styles	Wechsler et al., 2012	Divergent
Short Scale of Creative Self	Karwowski et al., 2018	Creativity is an important part of myself.

Note. Tests are ordered by publication year. For all these measures, raters express agreement with the descriptors on either a five-point Likert scale or a two-point (agree, disagree) scale. A few items were rephrased.

L-data in self-report is based on biographical information like "how often have you carried out this activity over the past ten years: wrote a blog entry?" (Diedrich et al., 2018). With this information, they try to assess if a specific situation has ever occurred in a person's life and how often. Table 6 summarizes everyday creativity measures. The Independent Activities Questionnaire (Schultz & Skager, 1963; Skager et al., 1965) described in Table 6 assesses the quantity and quality of previous creative accomplishments that participants listed in an open answer measure. The Creativity Scale for different domains, for example, is a domain-specific measure of creativity, although the authors discuss that general creativity is highly correlated with domain-specific creativity (Kaufman & Baer, 2004).

Table 6

Self-report (Accomplishments): Measures of Creative Activities.

Measure	Author	Example Items
Independent Activities Questionnaire	Schultz & Skager, 1963; Skager et al., 1965	Do you sometimes discuss books you have read with friends or family?
Creativity Scale for different Domains/ Creativity Domain Questionnaire	Kaufman & Baer, 2004; Kaufman et al., 2010	How creative are you in the area of crafts (for example, wood- working, sewing, repairing things, building things, cooking, etc.)?
Life-Space Questionnaire	Ivcevic & Mayer, 2009	Designed a video game.
Inventory of creative activities and achievements	Diedrich et al., 2018	How often have you written a blog entry over the past 10 years?

Note. Tests are ordered by publication year. The participant should indicate how often they have already participated in a listed behavior. Some items were rephrased.

Measures assessing previous creative achievements often require a listing of achievements in a given time-span. This can either include achievement checklists (e.g., philosophy of life changed; Torrance, 1969) in areas such as fine arts, crafts, literature, music, performing arts, and math-science (the Creative Behavior Inventory; revised by Dollinger, 2003), or can be based on open-ended answers (e.g., “Wrote a play that was publicly performed or a screenplay for a film that was publicly shown.”, Stricker et al., 2001). A more recent approach based on a Likert-scale is the Creative Achievement Questionnaire (Carson et al., 2005), including ten domains of creative achievements: visual arts, music, creative writing, dance, drama, architecture, humor, scientific discovery, invention, and culinary. The Inventory of Creative Activities and Achievements (Jauk et al., 2014; Diedrich et al., 2018) includes an activity scale reporting how often several activities were conducted within the last ten years (similar to Hocevar, 1979 and Carson et al., 2005). These scales are pursuing the same measurement intentions as the Biographical Inventory of Creative Behavior (“In the past

12 months have you designed and produces a textile product?"; Batey & Furnham, 2008). Only two measures have been applied in the school context: The Biographical Inventory for Students (Siegel, 1956) uses adjectives to assess how frequently children participate in creative activities. Measures that are important for the workforce have been mostly applied to identify creative personnel (research personnel, Buel, 1965; creative engineering personnel, Gough´s Adjective Checklist, McDermid, 1965; NASA scientists and engineers, Alpha Biographical Inventory, Institute for Behavioral Research in Creativity, 1968; Biographical Creativity Predictor for Scientific and Technical personnel, Erickson et al., 1970; Biographical Inventory Creativity, Schaefer & Anastasi, 1968). Additionally, the Achievement in Leadership questionnaire (e.g., Appointed to one or more offices; Holland & Nichols, 1964) and the Life Experience Inventory (Michael & Colson, 1979) include biographical data on creativity. The lifetime creativity scale is based on interview data and measures for real-life creativity at work and leisure. It is applied to assess the originality of creative achievements and their adaptation to reality (Richards et al., 1988).

Psychometric Shortcomings in Self-Reports

Self-report measures are susceptible to several sources of bias—social desirability, reference, and response style biases. In the assessment of creativity, individuals' introspection might be problematic (e.g., Baas et al., 2015). There has been much effort in the development of creativity self-report measures within the last few years (e.g., Forgeard & Kaufman, 2015; Kaufman, 2019; Silvia et al., 2012), and recent measures often show acceptable to good reliability (e.g., Revised Creative Domain Questionnaire: α = .82; the Biographical Inventory of Creative Behaviors α = .74; Silvia et al., 2012). However, reliability only reflects consistency in responding, and that consistency can be inflated by the consistency of biases. A way to get around this is by considering correlations with non-self-report measures of creativity. However, the convergent validity of self-report scales with other creativity indicators—such as divergent thinking—is alarmingly low for a wide range of measures (e.g., r = .16; Kandler et al., 2016). Besides, self-reports are prone to various self-report biases (e.g., social desirability, acquiescence, highly fakeable, etc.). As mentioned above, the self-report measures (Q-data and L-data) reported in our review require some amount of introspection of the individual. This can be problematic as individuals may understand the question, item, or response options differently (Weiss & Roberts, 2018) or show an incongruence between their beliefs regarding creativity and the actual scientific evidence (Baas et al., 2015). This can lead to low reliability of a self-report scale and low congruent validity with ability measures. This leads to the question of how well an individual (or others') can rate themselves (or others') on a self-report if researchers are still struggling to define and understand creativity. Another vital issue displayed in self-reports (especially the Q-data) is that the fixed choice questions lack flexibility and force people to respond. Many Q-data items are not in line with the bipartite nature of the definition of creativity, as they are very specific and often do not meet the definition.

Furthermore, these items lack a real test of the people's ability to be original and fluently produce novel things. The measures that are identified as Q-data are also not easy to distinguish from other close constructs. For example, there might be a substantial overlap between personality self-report and creativity self-report items. This overlap is further described in the discussion. Hence, practitioners must use these items carefully to avoid jingle-jangle fallacies. The measures, including L-data, also require introspection regarding an extended period, and some participants might not be able to report any creative achievements.

Constructs in Other-Reports

Creative attitudes, characteristics, and divergent thinking are constructs of other-reports (Table 2). Divergent thinking, or as Guilford (1950) labeled it, divergent production, is a cognitive process that produces original ideas in contrast to convergent thinking that results in a single correct answer. More details regarding divergent thinking are provided in the Constructs Section of Ability Tests. Creative attitudes and characteristics can be defined precisely as in self-report measures and are similar in what they measure, with the difference that they are based on significant others' ratings. Another construct assessed by experts (significant others') is ratings of archived products' originality and novelty from eminent persons. In this case, the measure is mostly based on products, actions, and past ideas.

Data Type and Scoring in Other-Reports

Other-reports are be based on peer or parent's reports, supervisor or teacher ratings, or experts judging the creative amount of a product, performance, or action. Hence, the data type includes subjective questionnaires and Life-data. Others' assessments based on Q-data describe the typical behavior of a target. At the same time, L-data consists of the ratings of maximal cognitive efforts or skill-related person's typical behavior. These assessments have a long history and are often based on the Consensual Assessment Technique (CAT; Amabile, 1982). The CAT is based on the idea that the experts consent on the creativity of artwork, creative stories, or research proposals and that this consent is the best measure of creativity (Baer & McKool, 2009). Typically, the participant is aware of the observation (e.g., the participant is applying for an award, attending a competition). This can contrast the others' assessment of typical behavior when the target is not aware of the observation (a supervisor rates the participant's overall creativity post-hoc or unobtrusively).

Other-Reports Measures

The measures we have identified that include others' divergent thinking ratings focus on children's creativity in terms of divergent production (see Table 7). Ratings include peer ratings of characteristics of creative children (Reid et al., 1959), teachers' ratings (Gifted and Talented Screening Form; Johnson, 1979), parents' ratings (Preschool and Kindergarten Interests Descriptors; Rimm, 1983), and caregivers' ratings (Creative Dramatics Test; Hensel, 1973). Another important branch of other-reports focuses on employee creativity (e.g., Ng & Feldman, 2012). A meta-analysis identified 86 studies, of which 60% included some supervisory ratings of creativity and five percent peer ratings. Even though such ratings are applied and of high value for companies, explicitly published measures for this purpose are sparse. Exemplary items that often have been applied in such studies are: (1) "Suggests many creative ideas that might improve working conditions.", (2) "Often comes up with creative solutions to problems at work.", (3) "Suggests new ways of performing work tasks.", and (4) "Is a good source of creative ideas." (e.g., Baer & Oldham, 2006; Madjar, Ortiz-Walters, 2008; Zhou & George, 2001).

Table 7

Other-report (Preferences): Divergent Thinking

Measure	Author	Example Item
Characteristics of Creative Children	Reid et al., 1959	My classmate has new ideas and new ways of doing things.
Creative Dramatics Test	Hensel, 1973	The child has creative thinking abilities.
Gifted and Talented Screening Form	Johnson, 1979	The child is talented in creativity.
Creativity checklist	Johnson, 1979	The child has a preference for complexity.
Preschool and Kindergarten Interests Descriptors	Rimm, 1983	My child tends to be very curious.
Leadership and Employee Creativity	Tierney et al., 1999	The employee took risks in terms of producing new ideas on the job.
Supervisor-ratings of Creativity	Zhou & George, 2001	Suggests many creative ideas that might improve working conditions.

Note. Tests are ordered by publication year. A few items were rephrased.

Creative attitudes and characteristics rated by others’ are also often focused on children. The Parental Evaluation of Children’s creativity (Runco, 1989) includes creative attitudes, intellect, and motivation. Either teachers or parents can complete the Creative Assessment Packet (Williams, 1993) (e.g., “My child has a vivid imagination.”). The Scales for Rating the Behavioral Characteristics of Superior Students (Renzulli et al., 2002; e.g., “The student demonstrates creative thinking about scientific topics.”) and the Gifted and Talented Evaluation Scales (Gillian et al., 1996; e.g., “The student demonstrates enthusiasm in discussions of scientific topics”) are based on teachers judgments such as the Ideal Pupil Checklist (Torrance, 1975) and the Creative Behavior Inventory (Kirschenbaum, 1989; e.g., “This child notices and remembers details.”)

The post-hoc evaluation and study of eminent people is one of the first attempts to measure and understand creativity and was applied before most other creativity measures. Analyzing the archival data of eminent people provides insight into the creative process. Most eminent people have produced numerous ideas and outcomes, with only rare breakthroughs sometimes driven by luck and chance (e.g., Pasteur or Nobel; Cropley, 2006). One early example is studying the thousand most eminent persons in history (Cattell, 1903). Others' have identified people who had produced highly creative and impactful outcomes in their field (Roe, 1951; Barron, 1969; Hall & MacKinnin, 1969; Helson & Crutchfield, 1970; Helson, 1971). Cattell´s (1903) study is based on rank-ordering the most eminent people based on the number of words written about the people in various biographical dictionaries. Based on this idea, the most creative architects, writers, biologists, and mathematicians were selected. This is also an example of the known-groups technique widely used in organizational psychology (Rhoades & Landy, 1973). Table 8 summarizes attempts to study the eminence of persons.

Table 8

Other-report (Accomplishments): Archival Data

Domain	Author	Aim of the study
Overall eminence	Cattell, 1903	Rank-ordering the 1,000 most eminent people.
Physical scientists	Roe, 1951	Rating the excellence of physical scientists on a three-point scale.
Leaders	Barron, 1969	Selecting and studying leaders in Irish Economic life.
Architects	Hall & MacKinnin, 1969	Identifying and studying the 40 most creative architects in the United States.
Math	Helson & Crutchfield, 1970; Helson, 1971	Identifying the most creative researchers in math and studying their achievements.
Engineering	Chakrabarti, 2013	Studying the innovations of engineering designers and innovators.
Sports	Martin & Cox, 2016	Analyzing the moves of a basketball player.

Note. Ordered by publication year. For a more detailed overview of the study of eminent people until 1979, see Hovecar (1979)

A further construct measures the originality and novelty of ideas and products. A product is original if it is uncommon and somehow novel (Maltzman, 1960). In work and life, others rate the originality and novelty of ideas generated (e.g., originality of ideas involving the US. Patent Index; Owens et al., 1957). Indices like patents and invention disclosure forms or number of research reports are rated to assess employee creativity (Tierney et al., 1999). Such indices (also including infrastructure or cultural environment) are also applied to rate the creativity of a state or country (e.g., creative economy; Correia, 2014)[1]. On the other hand, originality is often assessed in visual arts and poems (Cattell et al., 1918; Brittain & Beittel, 1964; Csikszentmihalyi & Getzels, 1970; Baer et al., 2004) or science (Harmon, 1963; Taylor et al., 1963). For example, competitions often include a judgment of originality and novelty (e.g., NYC radio contest: submit a humorous and original little green thing; Ward & Cox, 1974). Judgments of product originality and novelty can also take place in school (e.g., mathematics

[1] These indices are not further included as they focus on a whole country rather than an individual. For an overview see Correia (2014).

and literature, Foster, 1971; creativity in writing, Wallen & Stevenson, 1960). Product creativity in schools was shown to be highly correlated with intelligence (*r* = .57, *N* = 63) and school grades (*r* = .66, *N* = 63). Most of the above-described evaluations such as stories, art, poetry, and aesthetic products (Amabile, 1982; Horn & Salvendy, 2006; Cropley & Kaufman, 2012; Jeffries, 2017) are based on consensual assessment technique (Amabile, 1982; Cseh et al., 2016). Another recent scale is the Creative Solution Diagnosis Scale (Cropley & Kaufman, 2012) that is applied for judging artworks, cartoons, t-shirts, novel chairs, and advertisements (Besemer & O'Quin, 1986, 1987, 1999; White et al., 2002) and even mouse-trap designs. Further scales are the Creativity Product Inventory (Taylor & Sandler, 1972; Taylor, 1975), Creative Product Semantic Scale (O'Quin & Besemer, 1989, e.g., a product is surprising and germinal), Student Product Assessment Form (Renzulli & Reis, 1981), and the usability assessment in the area of technological and scientific products (Han et al., 2000; Cropley & Kaufman, 2012).

Psychometric Shortcomings in Other-Reports

Other-reports are common across student and class evaluation by teachers, peers or parents, and the workplace (e.g., supervisor-ratings). Ratings can be based on various events and situations, and the raters can have different expertise and biases. Drevdahl (1956) showed that teachers' nominations were not significantly different when the teachers were given a clearly defined concept of creativity and were clearly instructed. However, parents' ratings are often only weakly related to teachers' ratings of the students' creativity (Runco, 1989). This implies the importance of a standardized instruction, the test material (e.g., the granularity of the Likert-scale; Cseh & Jeffries, 2019) and domain expertise (e.g., Amabile, 1996), which is often not given. Concerning expertise, an important question is who an expert is and how many years of experience can someone claim to be an expert, and how many experts should be asked to gain a valid consensual judgment (Cseh & Jeffries, 2019).

In conclusion, all other-rating can be psychometrically problematic if the rater is not very familiar with the target or the rated construct. Meta-analytic studies provide evidence that the frequency of interaction with the target and especially interpersonal intimacy is necessary to substantially increase the accuracy of other-rating on personality (Connelly & Ones, 2010). Besides, the data of human raters are intrinsically nested. Such nestedness and the effects of the human coding should preferably be part of the statistical modeling, as, applied in specific item response theory models, that account for item and rater-characteristics (e.g., Robitzsch & Steinfeld, 2018). We assume that with an ambiguous construct like creativity, the rater must be familiar with the target to gain a reliable rating. Many studies in school and the workplace fail to require sufficient familiarity.

We conclude that either assessing eminent people by their fluency or the originality and novelty of products is a promising approach. In some settings, the construct's definition is not clear, and there are no guidelines for how to rate creativity. L-data ratings of creativity mostly include judges familiar with the rated object, but this approach is highly domain-specific and often does not meet the bipartite definition of creativity. Scientific achievements (e.g., patents) might be novel or original and useful. But usefulness may be a harder standard to meet when it comes to arts or music.

Constructs in Ability Test

The application of ability tests in assessing creativity arguably began as part of the United States Army Air Force Aviation Psychology Program Psychology (Guilford, 1947). It continued through to Guilford’s Aptitudes Project carried out at the University of Southern California (Guilford et al., 1952). Guilford’s 1950 Presidential Address to the American Psychological Association (Guilford, 1950) is often considered a turning point and strongly influenced the way creativity was perceived. Guilford argued that the study of creativity had been neglected. A systematic factor analytic investigation of primary creative abilities, such as the constructs of fluency, flexibility, and the production of novel ideas, would be useful. Wilson and colleagues (1954) administered 53 tests and ran a factor analytic study of creativity a few years later. This resulted in 46 different divergent thinking tests categorized by 14 constructs of creative thinking ability that are not further referred to in this review (a complete list of these measures is provided in the Ability Tests Measure section). Later, Guilford (1956, 1959, 1960, 1986) claimed that divergent thinking is comprised of fluency (quantity of adequate responses), flexibility (number of category switches within various answers), originality (classifying the rarity, cleverness, and originality of a response within the sample), and elaboration (amount of details given in an answer). Nevertheless, divergent thinking tests cannot be easily subdivided by fluency, flexibility, and originality. These aspects represent scoring methods, and many tests are scored on more than one dimension (e.g., originality and quantity of named things).

Besides divergent thinking, the other creativity construct in ability measures includes insight performance in creative problems (Polya, 1945). Insight describes the process of solving a problem (Mayer, 1995) and has often been related to major scientific discoveries and creativity (Cropley, 2006; Finke, 1995). Insight is only involved when solving novel and non-routine problems and includes a shift of perspective (Dow & Mayer, 2004). This shift of perspective and discovering new ideas are also referred to as the *Eureka* moment (Vallée-Tourangeau, 2018; Sprugnoli et al., 2017). Insight tasks are often employed as creativity tasks (e.g., Gibson et al., 2009; Martindale et al., 1984). They are meant to facilitate creative solutions because of their nature of linking remote ideas or words (Mednick & Mednick, 1967).

Data Type in Ability Tests

In most divergent thinking tasks, participants are asked to generate ideas or react to a spatial or verbal prompt (Kim, 2006). The tasks are used to gather objective test-data by subjects providing open-ended responses. The behavior measured here is maximal cognitive effort. Performance on insight tasks is usually expressed as the proximity to a pre-defined solution. Insight tasks can be verbal, mathematical, or spatial problems that are all featured by one correct answer. Hence, they are often scored as the participants´ proportional attainment of an objective response standard.

Ability Tests Measures

Many divergent thinking tests go back to the tests Wilson and colleagues have proposed (1954). The tests include verbal, numerical, or spatial content. Tests presented by Wilson included, for example, Brick Uses/Unusual Uses ("List different uses for a brick."), Consequences Test ("List consequences of certain changes."), and Plot Titles ("Write titles for story plots: fluency score/originality score."). For an exhaustive list of measures, please see the list provided in the OSF. In the following, we present divergent thinking tests developed after Wilson et al. (1954). As the construct of divergent thinking includes most tests, we present them by clustering them as conventional divergent thinking tests and tests for specific contexts (such as occupational, dramatic arts, etc.) and tests for children.

Table 9 summarizes the tests that include verbal and spatial stimuli and require participants' verbal and spatial production. In this table, new test versions are presented, whereas adaptions of the tests Wilson and colleagues (1954) proposed are not included. One very prominent divergent thinking test—also included in Table 9—is the Torrance Test of Creative Thinking (Torrance, 1966). The test includes verbal (e.g., “Improve products”) and spatial subtests (e.g., “What might this picture show?”), that are combined in one sum score. The test is based on fluency, flexibility, and originality aspects of creativity. It includes verbal subtests such as asking, guessing causes, guessing consequences, product improvement, unusual uses, unusual questions, and spatial subtests such as picture construction, picture completion, and lines and circles. The spatial test is assumed to be a culture-fair instrument as it is only based on minimal language. In a 50-year follow-up longitudinal study, the scores obtained in late 1950 were correlated to current personal and public achievement. It was found that these scores were moderately correlated with personal achievement but not with public achievement (Runco et al., 2010). The Educational Testing Service's (ETS) manual for the kit of reference tests for cognitive factors (French et al., 1963; Ekstrom et al., 1976) also included various divergent thinking tasks for flexibility, fluency (associational, expressional, ideational, word), and originality (including modifications of the tests proposed by Wilson et al., 1954; like the originality tests: plot titles, symbol production, and consequences).

Table 9

Ability Tests: Divergent Thinking Spatial and Verbal (after Wilson et al., 1954)

Measure	Author	Example Items
Consequences Task	Christensen et al., 1958	Imagine that people no longer needed to sleep. What would happen as a consequence?
Test of Creative Ability	Industrial Relations Center, 1959; Harris & Simberg, 1959	List all possible uses for an ordinary wire coat hanger.
Alternate Uses Task	Christensen et al., 1960	List all the different ways you could use a chair.
Line Meanings	Tagiuri, 1960	What does this line make you think of?
Minnesota Tests of Creative Thinking and Writing	Yamamoto, 1964	What questions can you think about the things you see in the picture?
Instances Task	Wallach & Kogan, 1965	Write down all of the unusual, creative, and uncommon instances of things that are round.
Wallach-Kogan Test	Wallach & Kogan, 1965	Tell me all the ways in which an apple and orange are alike. (Similarities)
Torrance Test of Creative Thinking	Torrance, 1966	What might this picture show?
Plot Titles	Berger & Guilford, 1969	Provide appropriate titles for a story.
Incomplete Figures	Torrance, 1969	Complete a given figure.
Picture construction from dots	Torrance, 1969	Construct a picture out of dots.
Comprehensive Ability Battery	Hakstian & Cattell, 1975	Generate ideas and alternative uses in applied arts, sales, marketing, consulting, and teaching.

Combining Objects	Ekstrom et al., 1976	Name two objects, when used together, fulfill a particular request.
Substitute Uses	Ekstrom et al., 1976	Think of a common object that could serve as a substitute for a given purpose.
Making Groups	Ekstrom et al., 1976	Combine given items in a group and provide a reason for grouping them.
Different Uses	Ekstrom et al., 1976	Provide other uses for a magazine.
Toothpicks Test	Ekstrom et al., 1976	Make different patterns of squares outlined by toothpicks.
Planning Patterns	Ekstrom et al., 1976	Fit figures onto a group of dots.
Storage Test	Ekstrom et al., 1976	Store objects in a given space.
Word Endings Test	Ekstrom et al., 1976	Write words that are ending in ATE.
Word Beginnings Test	Ekstrom et al., 1976	Think of words that are beginning with RE.
Word Beginning and Ending Tests	Ekstrom et al., 1976	Think of words beginning with S and ending with N.
Opposites Test	Ekstrom et al., 1976	Try to think about some words which mean the opposite of the word easy.
Controlled Associations Test	Ekstrom et al., 1976	Write as many synonyms for the word short.
Figure of Speech Test	Ekstrom et al., 1976	Try to think about words and phrases that complete the figure of speech: the jewels sparkled like…
Making Sentences	Ekstrom et al., 1976	Write a sentence contain words that begin with the letters E – R – T
Arranging Words	Ekstrom et al., 1976	Write sentences containing the words TAKE – FEW – LAND – LITTLE
Rewriting	Ekstrom et al., 1976	Rephrase the sentence “In response to the teacher´s question, a forest of hands shot up.”
Ornament Test	Ekstrom et al., 1976	Decorate a plain lampshade in different ways.
Elaboration Test	Ekstrom et al., 1976	Add details on a playing card.
Symbols Test	Ekstrom et al., 1976	Draw symbols that represent the word food.
Topics Test	Ekstrom et al., 1976	Write as many ideas you can think of about “A train Journey”.
Theme Test	Ekstrom et al., 1976	Write a paragraph about the theme “a tree”.
Thing Categories Test	Ekstrom et al., 1976	List things that are always red or that are red more often than any other color.
American Haiku	Amabile, 1985	Write a five-line, unrhymed poem, following a series of steps.
Emotional Consequences	Averill & Thomas-Knowles, 1991	What would happen if people would fall in love with a different person every day?
Word rearrangement test	Fillis & McAuley, 2000	Construct a story that contains as many of the presented words as possible.
Aurora Project	Chart et al., 2008	Imagine that the world changed so that almost everything can speak. Write a little conversation describing what the two things listed might say to each other if they could talk. Each thing must

		say at least one thing. Before you write what they will say, circle who is speaking.
Situational Originality	Antonietti et al., 2011	Choose three things, between the listed words to describe a tennis game.
Kaleidoscope Project	Sternberg, 2012	Share a one-minute video that says something about oneself.
Cued Creativity Verb Generation Task	Prabhakaran et al., 2014	Say a creative verb that is related to the noun scissors.

Note. Tests are ordered by publication year. A few items are rephrased.

In the following paragraph, we present several divergent thinking tests that were specifically adapted for the occupational context. The Owens' Creativity Task for Machine Design (Owens, 1960) includes tasks to design appropriate mechanisms for a functional machine. Further tests developed for engineering personnel are the Purdue Creativity Test (e.g., generate flexible ideas to solve job-related problems, Lawshe & Harris, 1960) and the Creative Engineering Design Assessment (Charyton & Merrill, 2009). Scientific creativity can be measured based on the formulating hypotheses test (Frederiksen et al., 1975). Each item consists of a depiction of research results, and the participant has to generate a possible hypothesis. The Scientific Word Association Test (Gough, 1975; Gough 1976) includes 100 scientific words, and the participant is asked to produce rare associations.

We include domain-specific tests used to assess creativity in art and music. The Denny-Ives Creativity Test (Rusch et al., 1964) assesses dramatic art based on fluency, redefinition, originality, and sensitivity. An example of measurement in the aesthetic domain is the Hall Mosaic Construction Test (Hall, 1972). Another pure spatial task is described by Niu and Sternberg (2001), including collage making and alien drawing to assess the cultural influences on artistic creativity between Chinese and American participants. Measures of divergent musical thinking are based on the idea of generating a response to a musical stimulus or problem (e.g., Measure of Creativity in Music and Sounds, Baltzer, 1988; Measures of Musical Divergent Production, Gorder, 1980; Richardson & Saffle, 1983;

Onomatopoeia and Images, Torrance et al., 1973; Thinking Creatively with Sounds and Words, Torrance et al., 1973, Houtz, 1985; Musical Expression Test, Barbot & Lubart, 2012). The Measure of creative Thinking in Music (Webster, 1994) measures flexibility by category switches in musical parameters (like dynamics) and originality by the statistical infrequency of a response. A specific test is the test of Musical Originality that includes the composition of creative musical pieces based on tonal repetitions and variances (Coffman, 1992).

Lastly, we describe tests for children that might be of special interest to caregivers and other practitioners. Tests suited for very young children are the Divergent Movement Ability Test (Cleland, 1994) that engages in motor creativity (e.g., new motor patterns) and Thinking Creativity in Action and Movement (Torrance & Gibbs, 1977). Other tests for children are the Starkweather Creativity Tests available for pre-school children (Starkweather, 1971), the Purdue Elementary Problem Solving Inventory (Feldhusen et al., 1972), the Gross Geometric Forms Test (Gross & Marsh, 1970), and the originality of forms test that consists of three-dimensional abstract forms (Starkweather, 1974). The Barron-Welsh Art Scale is a Freudian based assessment in which the participant is asked to draw a picture. The pictures are scored on various scales, including personal styles and social attitudes (Welsh & Barron, 1959). The Test for Creative Thinking-Drawing Production (Jellen & Urban, 1989) uses a set of eleven criteria to assess the creativity of a drawing produced by elementary school children. Other tests that have been applied in the school context are the Creative Reasoning Test (Doolittle, 1989) and the Evaluation of Potential Creativity (Lubart et al., 2011). The latter includes tasks similar to the early Guilfordian tests and is based on the idea that divergent-exploratory and convergent-integrated thinking is important for creative processes. A more recent approach for measuring scientific creativity in children is the Creative Scientific Ability Test (Ayas & Sak, 2014). The test measures fluency, flexibility, hypothesis testing, generation, and evidence evaluation in five subtests that are allocated to the fields of Biology, Interdisciplinary Science, Chemistry, Physics, and Ecology.

The second construct in ability tests includes creative insight problems. In classic insight problems, the participant is asked to provide one correct solution. Historically, Gestalt-psychological problems were used to study insight performance (e.g., Koehler, 1967). Insight has been traditionally studied with problems such as the *Nine-dot Problem*, the *Dunker Candle Task,* and the *The Triangle of Coins* (for an overview, see Chu & MacGregor, 2011). Besides these classic insight tasks, Chu and MacGregor (2011) also reviewed newer insight problems, such as Matchstick Arithmetic (Knoblich et al., 1999), Compound Remote Associates (Bowden & Jung-Beeman, 2003) or Remote Associate Tasks (Mednick & Mednick, 1967) or Rebus Puzzles (MacGregor & Cunningham, 2008). In the *Remote Associate Task* the participants are asked to produce a remote associate to a given word. Other typical insight problem tasks are anagrams (Novick & Sherman, 2003) and riddles (Luo & Niki, 2003) as a measure of word imagery representing verbal problems (Dewing & Hetherington, 1974). Word analogies have also been assessed together with different instructions (such as "Think more creatively about whether the four-word set constitutes a valid analogy.", Green et al., 2012). Another verbal task is the brainteaser (Sheth et al., 2009), while the binarized images (Giovannelli et al., 2010) are task-based on pictorial stimuli. All these tasks are also described in Sprugnoli and colleagues' (2017) and Chus and MacGregors' (2011) review of human performance on insight tasks. A further, more recent test is the Design Thinking Creativity Test (Hawthorne et al., 2016) that presents problem-solving skills within the twenty-first century and is based on design thinking principles. Here the individuals must develop a "useful and meaningful" design solution (Plattner et al., 2015).

Psychometric Shortcomings in Ability Tests

Divergent thinking tasks are the most prominent tasks in measuring creativity and the tasks that have arguably better validity, reliability, and predictive power for relevant outcomes. We now turn to some issues of direct measures of creativity based on task performance. One of the biggest challenges comes with divergent thinking scoring (e.g., Cseh & Jeffries, 2019). First, the scoring should be in line with the instruction. This implies that the participants should be only scored on the dimension that was instructed and measured (e.g., in a task where the participants were instructed to be fluent, their fluency should be scored, while tasks that ask for an original answer should only account for originality). Second, the free-response format requires some human judgments. The problems that human judges contribute were already discussed above (e.g., see Robitzsch & Steinfeld, 2018 for models that account for different rater characteristics). The next difficulties arise when it comes to the dimension (e.g., originality) of the score that is derived from the data (Reiter-Palmon et al., 2019). Fluency, mostly instructed and derived from divergent thinking tasks, is evaluated rather easily and straightforward by a count variable that considers all correct answers. Flexibility—evaluated less frequently—counts the number of diverse answers. Originality, however, is the most complex scoring dimension, and the proposed ways of deriving an originality score vary (Reiter-Palmon et al., 2019). An early approach suggests scoring the dimensions uncommonness, remoteness, and cleverness (Wilson et al., 1953). Hence, a very creative idea should be one that is not too frequently named (uncommon), distant from obvious solutions (remoteness), and includes imaginativeness, funniness, or cleverness (French et al., 1963). These dimensions are often aggregated to a single subjective score of originality (Hocevar, 1979; Forthmann et al., 2017). Forthmann and colleagues (2017) described modern approaches for scoring these three dimensions, including computerized frequency scorings, associational remoteness based on latent semantic analysis, and a snapshot scoring for cleverness based on three human raters. Studies show that the correlations between latent variables for the scored dimensions were small to medium-sized ($r = .24$ to $r = .50$). Nevertheless, because all three dimensions are rarely applied simultaneously, and the nature of aggregated scores is still unexplored, a simple aggregation over various divergent thinking indices is considered problematic (Runco & Acar, 2012). Simple frequency (uncommonness) scorings are often applied in the literature as they are more time-efficient (Cropley, 1967; Vernon, 1971), and the cost considerations might be a factor in determining how to best score such measures (e.g., Hargreaves & Bolton, 1972; Carroll, 1993). In sum, the scoring is challenging and must be planned and conducted very carefully (for an overview, see Reiter-Palmon et al., 2019). Fluency scores that are easy to derive can be criticized concerning their dependability on the ability to retrieve information (e.g., broad retrieval ability highly predicts creative fluency, Silvia et al., 2013). Flexibility scores are often based on rather arbitrary classification systems that have varying granularity of categories. Originality scores can be problematized with regard to the lack of generally recognized standards and protocols, and often enough because of low agreement between raters. Another general shortcoming that is not further related to the scoring issues described above is that most divergent thinking tasks are time-constrained and based on verbal or spatial expressions. This indicates that intelligence and speed are sources of variance that may not be intended or even task-relevant.

Insight tasks are easy to score, but there is a question about their relation to creativity. Insight tasks require a correct response, but there is no possibility of assessing a person's novelty or originality. It may be challenging to know what process the person was applying to solve the problem. The convergent nature of insight tasks—one veridical answer—questions their applicability of creativity tasks and their detachability of typical convergent thinking (aka intelligence) tests. In fact, famous insight tasks, such as the Remote Associate Task (Mednick & Mednick, 1967), are used to assess both convergent thinking and divergent thinking. Therefore, some researchers argue insight is an important part of creativity (e.g., Weiss et al., 2020), while others describe no correlations with creativity but rather with intelligence and motivation (e.g., Shaw & Conway, 1990).

Discussion

General Summary and Conclusion

In this study, we have summarized previous findings of the taxonomies of creativity measures. We have derived and applied an approach that considers and integrates previously-suggested attributes. The task attributes considered in this review are measurement approach (e.g., self-report), construct (e.g., creative interest), data type (e.g., Questionnaire-data), prototypical scoring (e.g., scale scores), and psychometric problems (e.g., missing introspection). We found that it was possible to reliably classify 213 measures retrieved from the literature by these attributes according to the eclectic taxonomy. By providing sample items, we hope that readers will find the taxonomy informative, useful, and easy to understand. Two essential aspects of evaluating the taxonomy include discussing the reliability and validity of the assignment provided in this paper and retrieving or suggesting hitherto unconsidered indicators for creativity measurement. In the following, we summarize the advantages of the proposed taxonomy before we address the above-stated issues of validity of creativity measures with regard to their assignment in the taxonomy. Next, the taxonomy application is further explained and discussed before we close by presenting perspectives for future measures, including ideas for new indicators of measurement.

The taxonomy should prove helpful to both practitioners and researchers for several reasons. First, the categorization of measures according to attributes is straight-forward, as we showed by nearly-perfect rater agreement between raters. Second, competing or complementary measures can be compared with respect to several attributes (e.g., construct) and facilitate researchers' and practitioners' decisions. Hints concerning issues with individual tasks such as psychometric shortcomings can be easily identified. Third, novel measures can also be classified easily into the taxonomy, and predictions concerning convergent and discriminant relations can be derived.

We suggest that the present taxonomy is generative, meaning that future measures can be subsumed within the proposed attributes. The taxonomy is pragmatic as it allows researchers and practitioners to identify measures suitable for specific target groups quickly (e.g., children) or purposes (e.g., creativity in the domain of music). We will now turn to discuss general recommendations derived from the taxonomy.

Construct Validity of Creativity Measures

Based on the vast range of measures embedded in our taxonomy, we must ask whether all these measures really assess a coherent dispositional trait best labeled as creativity. Similarly, it has to be asked if there is at least a coherence regarding the dispositional trait within the constructs. Situations in which different phenomena are subsumed below the same label are referred to as *jingle fallacies* (Thorndike, 1904). For example, self-reported creative activities and creative achievements are highly correlated with other creative achievement self-reports (r = .68; Jauk et al., 2014), which can be taken as proof of convergent validity. On the other hand, creative activities are only weakly predicted by fluency (β = .22) and originality (β =.25), and the relation between creative achievements and fluency/originality was mediated by creative activities (Jauk, et al., 2014). This study distinguishes between divergent thinking measures, activities, and achievements as they do *not* measure the same underlying construct. Nevertheless, the literature often presents results regarding *creativity* that do not distinguish between the substantial differences associated with the measurement approaches and constructs.

Assigned to our taxonomy, this situation would indicate that all measures proposed as creativity measures should be related to one another – i.e., they should demonstrate a positive manifold. If this positive manifold is not confirmed, we would, at the very least, recommend relabeling variables of interest to indicate that they do not measure the same underlying construct. Additionally, other things being equal, any two creativity measures should be correlated more highly with each other than with non-creativity indicators. For example, the correlation between two creativity indicators should, in general, be stronger than that between creativity and an intelligence task. This aspect of construct validity can be evaluated through Multi-trait-Multi-method (MTMM) analysis; Campbell & Fiske, 1959), which separates format or method (e.g., Likert-scale vs. performance test) from construct or trait (e.g., creativity vs. intelligence). The taxonomy we propose suggests factors that might be considered to conduct such analysis. Available findings help to address what might emerge from such analysis. In various studies, no substantial or high correlations have been found between measures of typical behavior (assessed with self-reports) and maximal cognitive effort (assessing abilities). The literature provides several examples of the little-to-no correlation between indicators (e.g., other´s reports that are only weekly related with one another and unrelated with self-report (Richards et al., 1964; Taylor et al., 1963), creativity self-reports that are only weakly or not related with divergent thinking (Clapham, 2004). These weak correlations endorse the conclusion that measures of both behaviors assess distinct features.

We recommend that 1) creativity indicators should only bear the same label if they are highly related to each other and 2) a strict and careful use of labels and theoretical approaches that further describe the similarities and differences between indicators. We believe that this application would prevent further research results published under the same label. In sum, avoiding or carefully considering *jingle fallacies* would shed more light on creativity as a construct and its true nature. Besides, we recommend the use of divergent thinking tests. As measured with divergent thinking, the importance of originality is already made clear by the definition of creativity. Originality is a key attribute of creative products and can be best measured in a test situation and product/idea evaluation.

This guideline becomes even more important when so-called proxy variables are applied to gather indicators of creativity. As outlined above, measuring creativity suffers from *jingle fallacies*. This issue becomes more problematic when applied measures tap into different constructs. Measuring creativity with different constructs comes along with several drawbacks. However, we recognize the importance of such proxy measures for research as these measures might help enlarge the nomological net of creativity. Besides, using proxy variables for creativity is reasonable when the correlation between the proxy variable and the targeted aspect of creativity is very strong. For example, the assessment of creativity based on openness measures has been justified by its positive relation with creativity (McCrae, 1987), although its correlation with creativity is far from unity. Similarly, conscientiousness has been used as a proxy for creativity based on its negative relation (George & Zhou, 2001). Several personality inventories have been applied to assess creativity even though we might classify them as creativity covariates/proxy variables (e.g., Cattell & Eber, 1964; Schmeidler 1965; Helson 1965; Heist & Yonge, 1968; Torrance & Khatena, 1970). More specific indicators such as risk-taking (Johnson, 1978), leadership skills (Karnes & Chauvin, 1985), and team climate (Anderson & West, 1996) have also been used as proxies. All these proxy variables capture dispositions more or less distinct from creativity, and they are often only weakly related to each other – both conceptually and empirically. Importantly, different forms of creativity would require different proxy variables. Taken together, we recommend assessing creativity based on preferably validated multivariate measures. However, multivariate studies allow considering different measures and even proxy variables; thus, they are worth studying to shed more light on the nomological net of creativity and related constructs.

Applying the Taxonomy

In sum, although the taxonomy classifies different measures based on attributes, this does not strongly imply converging validity evidence across methods for assessing creativity. This result is due to the heterogeneity in conceptualizations and measurement approaches. In other words, at this stage, it is not possible to guarantee the validity of individual indicators to deliver valid measures of the cell within which they are placed. Therefore, we recommend a careful selection of creativity measures in line with the definition and the intended construct.

Earlier recommendations for creativity tests lack a discussion of psychometric shortcomings. Some researchers recommend self-report measures, as they are simple to apply and easy to score (Hocevar, 1981; Forgeard & Kaufman, 2016; Kaufman, 2019). However, we believe that the lack of a real evaluation of originality and novelty paired with the lack of most individual's introspection ability when it comes to such a complex construct speaks against self-reports in creativity assessment. Other's report comes along with a similar problem when it comes to the familiarity not only with the construct but also with the target. We recommend that, if applying others' reports, the rating should be based on experts and should include accomplishments rather than preferences. All things considered, we recommend the application of divergent thinking for assessing an individual's creativity. Although several psychometric problems must be considered, divergent thinking measures deliver results showing the smallest discrepancy between the construct definition and the scope of measures. Future research should address the remaining deficiencies of divergent thinking measures – many of which relate to the scoring of divergent thinking tasks (e.g., Reiter-Palmon et al., 2019). These deficiencies include the dependency between originality and fluency (Barbot, 2018; Forthmann et al., 2018), stimulus dependencies (Barbot, 2018), and the relevancy of task instructions (Nusbaum et al., 2014).

These psychometric problems can be addressed by applying a broader range of tasks that are carefully and congruently instructed, analyzed, and scored (Reiter-Palmon et al., 2019). Multivariate studies including such a broad task range (e.g., the comprehensive test battery for measuring creativity rCAB offers a selection of divergent thinking tasks also complemented by experimental approaches (3-D divergent thinking), self-reports, and even other's reports) can lead to better generalizability and hence foster the understanding of individual differences in creativity (rCAB, 2011). Test development can also benefit from interdisciplinary work, including neuropsychological research (Benedek et al., 2019) and trials (such as the Multi-Trial Creative Ideation; Barbot, 2018), industrial psychology work on inventions and patents, and research on scoring verbal divergent thinking tasks based on the the computer linguistic approaches (Weiss & Wilhelm, 2020).

In sum, although divergent thinking tests are the most widely applied indicators, there has generally been little progress in divergent thinking test development since the 1960s. The plot titles test (from Guilford´s test battery)—for example—can serve as an originality indicator but should be combined with other divergent thinking tests for fluency and flexibility (Runco et al., 2016). These tests should then be carefully instructed and scored as the simple scoring of fluency comes with a loss of information.

We believe our taxonomy provides a useful guide for researchers and practitioners to choose a psychometrically sound creativity test for their specific purpose. Besides considering all dimensions of the taxonomy: measurement approach (e.g., the assessment of creative ability), construct (e.g., insight), data type (e.g., T-data), and prototypical scoring (e.g., human coding), we also suggest the consideration of the Cohen's kappa for selecting a test in a specific construct.

Perspectives in Future Creativity Research

There is a variety of prospects for creativity research. Here, we want to focus on two avenues: a) enlarging the nomological net of creativity and proximal constructs, and b) providing inspiration for hitherto not further suggested creativity measures. First, creativity is studied for over a century, but there is still a surprising gap regarding multivariate studies exploring the validity and relationship between the measurement approaches, including related constructs. For example, only a few studies include achievement measures and self-reports of creative interest and divergent thinking measures of fluency and originality (e.g., Jauk et al., 2014). Nevertheless, such extensive multivariate studies are necessary to improve our understanding of what creativity is and its relation to neighboring constructs/proxies. To better understand what creativity is, it is also worth studying if creativity should be further differentiated. This endeavor can be pursued by studying constructs that are supposedly highly similar—measurement wise and theoretically—to creativity. One example of such a construct is emotional creativity that should be further studied in relation to and differentiation from creativity (e.g., Ivcevic et al., 2007).

Concerning the second avenue, we need to acknowledge that especially divergent thinking tasks have not improved much since they were first proposed, even though they are a promising candidate for implementing new technologies and psychometric procedures. A lot of currently applied divergent thinking tasks are just alterations of tasks that were originally suggested in the 1950s (e.g., Wilson et al., 1954). Despite the remarkable developments in several technologies, the literature does not offer many new approaches to pioneering ability tests (either divergent thinking or problem-solving). However, such technologies, for example, virtual realities, provide researchers with the opportunity to measure the creative ability within a context that can be closer to real-life settings or detailed and contextualized fantasy worlds than standard divergent thinking measures can be. Virtual reality offers the chance to perform creatively and interact with other players within a pre-specified setting. We argue that such actions and interactions in a gamified way can depict the individuals' divergent thinking and group dynamics that might foster or hinder any creative behavior or outcome. Apart from that, the data that can be derived from such settings might offer other intriguing ways of scorings that are not, or only partly, dependable on human judges. Taken together, we believe that the development of divergent thinking settings (for individuals and groups) within such technologies would greatly further the understanding of creativity.

References

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

Alfonso-Benlliure, V., Meléndez, J. C., & García-Ballesteros, M. (2013). Evaluation of a creativity intervention program for preschoolers. *Thinking Skills and Creativity, 10*, 112-120. https://doi.org/10.1016/j.tsc.2013.07.005

Allport, G. W., Vernon, P. E., & Lindzey, G. (1960). *Study of values.* Oxford, England: Houghton Mifflin.

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology, 43*, 997-1013. http://dx.doi.org/10.1037/0022-3514.43.5.997

Amabile, T. (1996). *Creativity and innovation in organizations* [Harvard Business School Background Note (pp. 396-239)]. Boston: Harvard Business School.

Amabile, T. M. (1985). Motivation and creativity: Effects of motivational orientation on creative writers. *Journal of personality and social psychology, 48*, 393-399. https://doi.org/10.1037/0022-3514.48.2.393

Anderson, N., & West, M. A. (1996). The Team Climate Inventory: Development of the TCI and its applications in teambuilding for innovativeness. *European Journal of work and organizational psychology, 5*, 53-66. https://doi.org/10.1080/13594329608414840

Antonietti, A., Colombo, B., & Pizzingrilli, P. (2011). The WCR model of creativity. From concept to application. *The Open Education Journal, 4*, 80-89. https://doi.org/10.2174/1874920801104010080

Auzmendi, E., Villa, A., & Abedi, J. (1996). Reliability and validity of a newly constructed multiple-choice creativity instrument. *Creativity Research Journal, 9,* 89-95. https://doi.org/10.1207/s15326934crj0901_8

Ayas, M. B., & Sak, U. (2014). Objective measure of scientific creativity: Psychometric validity of the Creative Scientific Ability Test. *Thinking Skills and Creativity, 13*, 195-205. https://doi.org/10.1016/j.tsc.2014.06.001

Averill, J. R. (1999). Individual differences in emotional creativity: Structure and correlates. *Journal of Personality*, *67*, 331-371. https://doi.org/10.1111/1467-6494.00058

Averill, J. R., & Thomas-Knowles, C. (1991). Emotional creativity. In K. T. Strongman (Ed.), *International review of studies on emotion.* London: Wiley.

Baas, M., Koch, S., Nijstad, B. A., & De Dreu, C. K. (2015). Conceiving creativity: The nature and consequences of laypeople's beliefs about the realization of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(3), 340-354. https://doi.org/10.1037/a0039420

Baer, J., Kaufman, J. C., & Gentile, C. A. (2004). Extension of the consensual assessment technique to nonparallel creative products. *Creativity Research Journal, 16,* 113-117. https://doi.org/10.1207/s15326934crj1601_11

Baer, M., & Oldham, G. R. (2006). The curvilinear relation between experienced creative time pressure and creativity: moderating effects of openness to experience and support for creativity. *Journal of Applied Psychology*, *91*(4), 963-970. https://doi.org/10.1037/0021-9010.91.4.963

Baer, J., & McKool, S. S. (2009). Assessing creativity using the consensual assessment technique. In *Handbook of research on assessment technologies, methods, and applications in higher education* (pp. 65-77). IGI Global. https://doi.org/10.4018/978-1-60566-667-9.ch00

Baltzer, S. (1988). A validation study of a measure of musical creativity. *Journal of Research in Music Education, 36,* 232-249. https://doi.org/10.2307/3344876

Barbot, B., & Lubart, T. (2012). Creative Thinking in Music: Its Nature and Assessment Through Musical Exploratory Behaviors. *Psychology of Aesthetics, Creativity, and the Arts 6,* 231–242. https://doi.org/10.1037/a0027307

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. http://dx.doi.org/10.1037/aca0000233

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

Barron, F. (1969). *Creative person and creative process*. New York: Holt, Rienhart & Winston.

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

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., & Furnham, A. (2008). The relationship between measures of creativity and schizotypy. *Personality and Individual Differences, 45*, 816-821. https://doi.org/10.1016/j.paid.2008.08.014

Benedek, M., Christensen, A. P., Fink, A., & Beaty, R. E. (2019). Creativity assessment in neuroscience research. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 218-226. https://doi.org/10.1037/aca0000215

Berg, H., Taatila, V., & Volkmann, C. (2012). Fostering creativity–a holistic framework for teaching creativity. *Development and Learning in Organizations: An International Journal, 26*, 5-8. https://doi.org/10.1108/14777281211272242

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

Besemer, S. P., & O'Quin, K. (1999). Confirming the three-factor creative product analysis matrix model in an American sample. *Creativity Research Journal*, *12*, 287-296. https://doi.org/10.1207/s15326934crj1204_6

Besemer, S., & O'Quin, K. (1986). Analyzing creative products: Refinement and test of a judging instrument. *The Journal of Creative Behavior*, *20*, 115-126. https://doi.org/10.1002/j.2162-6057.1986.tb00426.x

Besemer, S. P., & O'Quin, K. (1987). Creative product analysis: Testing a model by developing a judging instrument. *Frontiers of Creativity Research: Beyond the basics*, 367-389.

Bowden, E. M., Jung-Beeman, M., & Beeman, M. (2003). One hundred forty-four compound remote associate problems: Short insight-like problems with one-word solutions. *Behavioral Research, Methods, Instruments, and Computers, 35,* 634-639.

Brittain, W. L., & Beittel, K. R. (1964). A study of some tests of creativity in relationship to performances in the visual arts. *Studies in Art Education, 2*, 54-65. https://doi.org/10.2307/1319647

Buel, W. D. (1965). Biographical data and the identification of creative research personnel. *Journal of Applied Psychology, 49,* 318-321. https://doi.org/10.1037/h0022518

Bughin, J., Hazan, E., Lund, S., Dahlström, P., Wiesinger, A., & Subramaniam, A. (2018, May 23). *Skill shift: Automation and the future of the workforce*. https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce

Bull, K. S., & Davis, G. A. (1980). Evaluating creative potential using the statement of past creative activities. *The Journal of Creative Behavior, 14,* 249-257. https://doi.org/10.1002/j.2162-6057.1980.tb00256.x

Byrd, R. E. (1986). *Creativity and risk-taking.* San Diego, CA: Pfeiffer International Publishers.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge: Cambridge University Press. https://doi.org/10.1080/00140139508925174

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

Cattell, R. B. (1958). What is" objective" in" objective personality tests*?". Journal of Counseling Psychology, 5*, 285-289. https://doi.org10.1037/h0046268

Cattell, J. M. (1903). A statistical study of eminent men. *Popular Science Monthly, 62,* 359-377.

Cattell, R. B., & Eber, H. W. (1964). *Sixteen personality factor questionnaire.* Institute for personality and ability testing.

Cattell, J. M., Glascock, J., & Washburn, M. F. (1918). Experiments on a possible test of aesthetic judgment of pictures. *The American Journal of Psychology*, *29,* 333-336. https://doi.org/10.2307/1414125

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin, 56*, 81-105. https://doi.org/10.1037/h0046016

Chakrabarti, A. (2013). Understanding influences on engineering creativity and innovation: a biographical study of 12 outstanding engineering designers and innovators. *International Journal of Design Creativity and Innovation, 1*, 56-68. https://doi.org/10.1080/21650349.2013.754643

Chart, H., Grigorenko, E. L., & Sternberg, R. J. (2008). Identification: The Aurora Battery. In J. A. Plucker & C. M. Callahan (Eds.), *Critical issues and practices in gifted education: What the research says* (pp. 281-301). Waco, TX, US: Prufrock Press.

Charyton, C., & Merrill, J. A. (2009). Assessing general creativity and creative engineering design in first year engineering students. *Journal of Engineering Education, 98*, 145-156. https://doi.org/10.1002/j.2168-9830.2009.tb01013.x

Christensen, P. R., Guilford, J. P., Merrifield, P. R., & Wilson, R. C. (1960). *Alternate uses.* Beverly Hills: Sheridan Psychological Services.

Christensen, P. R., Merrifield, P. R., & Guilford, J. P. (1958). *Consequences*. Consulting Psychologists Press.

Chu, Y., & MacGregor, J. N. (2011). Human performance on insight problem solving: A review. *The Journal of Problem Solving, 3*, 6-38. https://doi.org/10.7771/1932-6246.1094

Clapham, M. M. (2004). The convergent validity of the Torrance Tests of Creative Thinking and creativity interest inventories. *Educational and Psychological Measurement, 64*, 828-841. https://doi.org/10.1177/0013164404263883

Cleland, F. E. (1994). Young children's divergent movement ability: Study II. *Journal of teaching in Physical Education, 13*, 228-241. https://doi.org/10.1123/jtpe.13.3.228

Coffman, D. D. (1992). Measuring musical originality using information theory. *Psychology of Music, 20,* 154-161. https://doi.org/10.1177/0305735692202005

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46. https://doi.org/10.1177/001316446002000104

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*, 1092-1122. https://doi.org/10.1037/a0021212

Correia, C. M. (2014). Measuring creativity in the EU member states. *Investigaciones regionales: Journal of Regional Research, 30*, 7-26.

Craft, A. (2005). *Creativity in schools: Tensions and dilemmas*. London: Routledge. https://doi.org/10.4324/9780203357965

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

Cropley, D. H., & Kaufman, J. C. (2012). Measuring functional creativity: Non-expert raters and the Creative Solution Diagnosis Scale. *The Journal of Creative Behavior, 46*, 119-137. https://doi.org/10.1002/jocb.9

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

Cseh, G. M., Jeffries, K. K., Lochrie, M., Egglestone, P., & Beattie, A. A. (2016). A DigitalCAT: A Fusion of Creativity Assessment Theory and HCI. In *Proceedings of the 30th International BCS Human Computer Interaction Conference: Companion Volume* (p. 25). BCS Learning & Developement Ltd.

Cseh, G. M., & Jeffries, K. K. (2019). A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 159-166. https://doi.org/10.1037/aca0000220

Csikszentmihalyi, M., & Getzels, J. W. (1970). Concern for discovery: An attitudinal component of creative production. *Journal of Personality, 38*, 91-105. https://doi.org/10.1111/j.1467-6494.1970.tb00639.x

Davis, G. A., & Rimm, S. (1982). Group inventory for finding interests:(GIFFI) I and II: Instruments for identifying creative potential in the junior and senior high school. *The Journal of Creative Behavior*, *16*, 50-57. https://doi.org/10.1002/j.2162-6057.1982.tb00321.x

Davis, G. A. (1999). Barriers to creativity and creative attitudes. *Encyclopedia of creativity*, *1*, 165-174.

Dawson, V. L., D'Andrea, T., Affinito, R., & Westby, E. L. (1999). Predicting creative behavior: A reexamination of the divergence between traditional and teacher-defined concepts of creativity. *Creativity Research Journal, 12*, 57-66. https://doi.org/10.1207/s15326934crj1201_7

Dewing, K., & Hetherington, P. (1974). Anagram solving as a function of word imagery. *Journal of Experimental Psychology*, *102*, 764-767. https://doi.org/10.1037/h0036339

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

Dollinger, S. J. (2003). Need for uniqueness, need for cognition, and creativity. *The Journal of Creative Behavior*, *37*, 99-116. https://doi.org/10.1002/j.2162-6057.2003.tb00828.x

Domino, G. (1994). Assessment of creativity using the ACL: A comparison of four scales. *Creativity Research Journal, 7*, 21-23. https://doi.org/10.1080/10400419409534506

Doolittle, J. H. (1989). *The Creative Reasoning Test*. Midwest Publications.

Dow, G. T., & Mayer, R. E. (2004). Teaching students to solve insight problems: Evidence for domain specificity in creativity training. *Creativity Research Journal, 16*, 389-398. https://doi.org/10.1080/10400410409534550

Drevdahl, J. E. (1956). Factors of importance for creativity. *Journal of Clinical Psychology, 12*, 21-26. https://doi.org/10.1002/1097-4679(195601)12:1<21::AID-JCLP2270120104>3.0.CO;2-S

Durmysheva, Y., & Kozbelt, A. (2010). The creative approach questionnaire: Operationalizing Galenson's finder-seeker typology in a non-expert sample. *The International Journal of Creativity & Problem Solving, 20*, 35-55.

Ekstrom, R. B. French, JW, Harman, H., & Derman, D. (1976). *Kit of factor-referenced cognitive tests (rev. ed.)*. Educational Testing Service.

Epstein, R., Schmidt, S. M., & Warfel, R. (2008). Measuring and training creativity competencies: Validation of a new test. *Creativity Research Journal, 20*, 7-12. https://doi.org/10.1080/10400410701839876

Erickson, C., Gantz, B. S., & Stephenson, R. W. (1970). Logical and construct validation of a short-form biographical inventory predictor of scientific creativity. In *Proceedings of the Annual Convention of the American Psychological Association*. American Psychological Association.

Eysenck, H. J. (1996). *The measurement of creativity.* In M. A. Boden (Ed.), Dimensions of creativity (pp. 199–242). Cambridge, MA: MIT Press.

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

Feldhusen, J. F., Houtz, J. C., & Ringenbach, S. (1972). The Purdue elementary problem-solving inventory. *Psychological Reports, 31*, 891-901. https://doi.org/10.2466/pr0.1972.31.3.891

Fillis, I., & McAuley, A. (2000). Modeling and measuring creativity at the interface. *Journal of Marketing Theory and Practice, 8*, 8-17. https://doi.org/10.1080/10696679.2000.11501864

Finke, R. A. (1995). Creative insight and preinventive forms. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 255-280). Cambridge, MA, US: The MIT Press.

Forgeard, M. J., & Kaufman, J. C. (2016). Who cares about imagination, creativity, and innovation, and why? A review. *Psychology of Aesthetics, Creativity, and the Arts, 10*(3), 250-269. https://doi.org/10.1037/aca0000042

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.2019.04.0

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*. https://doi.org/10.1037/aca0000196

Foster, J. (1971). *Creativity and the teacher*. Macmillan.

Frederiksen, N., Evans, F. R., & Ward, W. C. (1975). Development of provisional criteria for the study of scientific creativity. *Gifted Child Quarterly*, *19*, 60-65. https://doi.org/10.1177/001698627501900110

French, J. W., Price, L. A., & Ekstrom, R. B. (1963). *Manual for kit of reference tests for cognitive factors.* Educational Testing Service. https://doi.org/10.21236/ad0410915

Fricke, B. G. (1965). *Opinion, attitude and interest survey handbook: A guide to personality and interest measurement.* OAIS Testing Program.

Gabora, L., & Kaufman, S. B. (2010). Evolutionary approaches to creativity. In J. Kaufman & R. Sternberg (Eds.),*The Cambridge handbook of creativity* (pp. 279-300). Cambridge, UK: Cambridge University Press.

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

George, J. M., & Zhou, J. (2001). When openness to experience and conscientiousness are related to creative behavior: an interactional approach. *Journal of Applied Psychology, 86*, 513-524. https://doi.org/10.1037/0021-9010.86.3.513

Gibson, C., Folley, B. S., & Park, S. (2009). Enhanced divergent thinking and creativity in musicians: A behavioral and near-infrared spectroscopy study. *Brain and Cognition, 69*, 162-169. https://doi.org/10.1016/j.bandc.2008.07.009

Gillian, J. E., Carpenter, B. O., & Christensen, J. R. (1996). Gifted and Talented Evaluation Scales (GATES). *Traducción Madrid: PSYMTE.*

Giovannelli, F., Silingardi, D., Borgheresi, A., Feurra, M., Amati, G., Pizzorusso, T., ... & Cincotta, M. (2010). Involvement of the parietal cortex in perceptual learning (Eureka effect): an interference approach using rTMS. *Neuropsychologia, 48*, 1807-1812. https://doi.org/10.1016/j.neuropsychologia.2010.02.031

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96. https://doi.org/10.1016/j.jrp.2005.08.007

Gorder, W. D. (1980). Divergent production abilities as constructs of musical creativity. *Journal of Research in Music Education*, *28*, 34-42. https://doi.org/10.2307/3345051

Gough, H. G. (1975). A new scientific uses test and its relationship to creativity in research. *The Journal of Creative Behavior*, *9*, 245-252. https://doi.org/10.1002/j.2162-6057.1975.tb00573.x

Gough, H. G. (1979). A creative personality scale for the adjective check list. *Journal of Personality and Social Psychology*, *37*, 1398-1405. https://doi.org/10.1037/0022-3514.37.8.1398

Gough, H. G. (1992). Assessment of creative potential in psychology and the development of a creative temperament scale for the CPI. In *Advances in Psychological Assessment*, 225-257. Boston, MA: Springer. https://doi.org/10.1007/978-1-4757-9101-3_8

Gough, H. G. (1976). Studying creativity by means of word association tests. *Journal of Applied Psychology, 61*, 348-353. http://dx.doi.org/10.1037/0021-9010.61.3.348

Gough, H. G., & Heilbrun, A. B. (1983). *The adjective check list*. Consulting Psychologists Press.

Gray, K., & Koncz, A. (2017, November). *The key attributes employers seek on students´ resumes.* https://www.naceweb.org/about-us/press/2017/the-key-attributes-employers-seek-on-students-resumes/.

Green, A. E., Cohen, M. S., Kim, J. U., & Gray, J. R. (2012). An explicit cue improves creative analogical reasoning. *Intelligence*, *40*, 598-603. https://doi.org/10.1016/j.intell.2012.08.005

Gross, R. B., & Marsh, M. (1970). An instrument for measuring creativity in young children: The Gross Geometric Forms. *Developmental Psychology, 3*, 267. https://doi.org/10.1037/h0029585

Guilford, J. P. (Ed.). (1947). *Army Air Forces Aviation Psychology Program research reports. Printed classification tests: Report no. 5* (J. I. Lacey, Collaborator). Washington, DC, US: US Government Printing Office. http://dx.doi.org/10.1037/14946-000

Guilford, J. P. (1960). Basic conceptual problems in the psychology of thinking. *Annals of the New York Academy of Sciences*, *91*, 6-21. https://doi.org/10.1111/j.1749-6632.1961.tb50913.x

Guilford, J. P. (1986). *Creative talents: Their nature, uses and development*. Buffalo, NY: Bearly Limited.

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

Guilford, J. P. (1963). *Guilford-Zimmerman Interest Inventory*. Beverley Hills, CA: Sheridan Psychological Services.

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

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

Guilford, J. P. (1959). Three faces of intellect. *American Psychologist*, *14*, 469-479. https://doi.org/10.1037/h0046827

Guilford, J. P., Wilson, R. C., & Christensen, P. R. (1952). *A Factor-analytic Study of Creative Thinking: Administration of Tests and Analysis of Results. II*. University of Southern California.

Hakstian, A. R., & Cattell, R. B. (1975). *CAB (comprehensive Ability Battery)*. Institute for Personality and Ability Testing.

Hall, W. B. (1972). A technique for assessing aesthetic predispositions: Mosaic construction test. *The Journal of Creative Behavior*, *6*, 225-235. https://doi.org/10.1002/j.2162-6057.1972.tb00935.x

Hall, W. B., & MacKinnin, D. W. (1969). Personality inventory correlates of creativity among architects. *Journal of Applied Psychology*, *53*, 322-326. https://doi.org/10.1037/h0027811

Han, S. H., Yun, M. H., Kim, K. J., & Kwahk, J. (2000). Evaluation of product usability: development and validation of usability dimensions and design elements based on empirical models. *International Journal of Industrial Ergonomics*, *26*, 477-488. https://doi.org/10.1016/s0169-8141(00)00019-6

Harmon, L. R. (1963). The development of a criterion of scientific competence. *Scientific creativity: its recognition and development* (p. 44-52). New York: Wiley.

Hargreaves, D. J., & Bolton, N. (1972). Selecting creativity tests for use in research. *British Journal of Psychology*, *63*(3), 451-462. https://doi.org/10.1111/j.2044-8295.1972.tb01295.x

Harris, R. H., & Simberg, A. L. (1959). *AC test of creative ability.* University of Chicago, industrial relations Center.

Hawthorne, G., Saggar, M., Quintin, E. M., Bott, N., Keinitz, E., Liu, N., …, & Reiss, A. L. (2016). Designing a Creativity Assessment Tool for the Twenty-First Century: Preliminary Results and Insights from Developing a Design-Thinking Based Assessment of Creative Capacity. In *Design Thinking Research* (pp.111-123). https://doi.org/10.1007/978-3-319-19641-1_8

Heist, P., & Yonge, G. D. (1968). *Omnibus personality inventory*. Psychological Corporation.

Helson, R. (1965). Childhood interest clusters related to creativity in women. *Journal of Consulting Psychology, 29,* 352-361. https://doi.org/10.1037/h0022401

Helson, R. (1971). Women mathematicians and the creative personality. *Journal of Consulting and Clinical Psychology, 36,* 210-220. https://doi.org/10.1037/h0030728

Helson, R., & Crutchfield, R. S. (1970). Mathematicians: The creative researcher and the average PhD. *Journal of consulting and clinical psychology, 34,* 250-257. https://doi.org/10.1037/h0029003

Hensel, N. H. (1973). The development, implementation, and evaluation of a creative dramatics program for kindergarten children (doctoral dissertation University of Georgia, 1973), *Dissertation Abstracts International, 34, 4562A. University Microfilms* No. 74-4816.

Hocevar, D. (1981). Measurement of creativity: Review and critique. *Journal of Personality assessment, 45,* 450-464. https://doi.org/10.1207/s15327752jpa4505_1

Hocevar, D. (1979). *The Development of the Creative Behavior Inventory (CBI).*

Hofstee, W. K., De Raad, B., & Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of personality and social psychology, 63,* 146-163. https://doi.org/10.1037//0022-3514.63.1.146

Hogan, R., & Hogan, J. (1992). Hogan Personality Inventory. *PsycTESTS Dataset.* https://doi.org/doi:10.1037/t02029-000.

Holland, J. L., & Nichols, R. C. (1964). Prediction of academic and extra-curricular achievement in college. *Journal of Educational Psychology, 55,* 55-65. https://doi.org/10.1037/h0047977

Holland, J. L., Fritzsche, B. A., & Powell, A. B. (1997). *The Self-directed Search (SDS): Technical Manual*. Psychological Assessment Resources.

Horn, D., & Salvendy, G. (2006). Consumer-based assessment of product creativity: A review and reappraisal. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *16*, 155-175. https://doi.org/10.1002/hfm.20047

Houtz, J. C. (1985). Review of Thinking Creatively with Sounds and Words. *Test Critiques*, *4*, 666-672.

Industrial Relations Center. (1959). *AC Test of Creative Ability: Test administration manual*. Chicago: Industrial Relations Center, University of Chicago.

Institute for Behavioral Research in Creativity. (1968). *Alpha Biographical Inventory*. Greensboro, North Carolina: Prediction Press.

Ivcevic, Z., Brackett, M. A., & Mayer, J. D. (2007). Emotional intelligence and emotional creativity. *Journal of Personality*, *75(*2), 199-236. https://doi.org/10.1111/j.1467-6494.2007.00437.x

Ivcevic, Z., & Mayer, J. D. (2009). Mapping dimensions of creativity in the life-space. *Creativity Research Journal*, *21*(2-3), 152-165. https://doi.org/10.1080/10400410902855259

Jackson, D. N. (1994). *French Jackson Personality Inventory–Revised manual*. Prot Huron: Sigma Assessment Systems

Jauk, E., Benedek, M., & Neubauer, A. C. (2014). The road to creative achievement: A latent variable model of ability and personality predictors. *European Journal of Personality, 28*, 95-105. https://doi.org/10.1016/j.paid.2013.07.129

Jeffries, K. K. (2017). A CAT with caveats: is the Consensual Assessment Technique a reliable measure of graphic design creativity?. *International Journal of Design Creativity and Innovation*, *5*, 16-28. https://doi.org/10.1080/21650349.2015.1084893

Jellen, H. G., & Urban, K. K. (1989). Assessing creative potential world-wide: The first cross-cultural application of the Test for Creative Thinking—Drawing Production (TCT-DP). *Gifted Education International, 6*, 78-86. https://doi.org/10.1177/026142948900600204

Johnson, D. L. (1978). *Inventory of Individually Perceived Group Cohesiveness (IIPGC).*

Johnson, D. L. (1979). *Creativity checklist*. Stoelting.

Johnson, D. L. (1979). *Gifted and Talented Screening Form*. Stoelting.

Johnson, J. A. (1994). Clarification of factor five with the help of the AB5C model. *European Journal of Personality, 8*, 311-334. https://doi.org/10.1002/per.2410080408

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*(2), 230-249. https://doi.org/10.1037/pspp0000087

Karnes, F. A., & Chauvin, J. C. (1985). *Leadership skills inventory*. DOK Publishers.

Karwowski, M., Lebuda, I., & Wiśniewska, E. (2018). Measuring creative self-efficacy and creative personal identity. *The International Journal of Creativity & Problem Solving, 28*(1), 45–57.

Kaufman, J. C. (2012). Counting the muses: development of the Kaufman domains of creativity scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts, 6*, 298-308. http://dx.doi.org/10.1037/a0029751

Kaufman, J. C., & Baer, J. (2004). Sure, I'm creative—but not in mathematics!: Self-reported creativity in diverse domains. *Empirical studies of the Arts, 22*, 143-155. https://doi.org/10.2190/26hq-vhe8-gtln-bjjm

Kaufman, J. C., & Plucker, J. A. (2011). *Intelligence and creativity.* In R. J. Sternberg & S. B. Kaufman (Eds.), *Cambridge handbooks in psychology*. *The Cambridge handbook of intelligence* (p. 771–783). Cambridge University Press. https://doi.org/10.1017/CBO9780511977244.039

Kaufman, J. C. (2019). Self-assessments of creativity: Not ideal, but better than you think. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 187-192. https://doi.org/10.1037/aca0000217

Kazdin, A. E., (2000). *Encyclopedia of psychology*. Washington, DC: American Psychological Association.

Kelley, T. L., (1927). *Interpretation of educational measurements*. Macmillan.

Khatena, J., & Morse, D. T. (1994). *Khatena-Morse Multitalent Perception Inventory: Scoring and Administration Guide: Norms-technical Manual.* Scholastic Testing Service, Incorporated.

Khatena, J., & Torrance, E. P. (1976). *Khatena-Torrance creative perception inventory*. Stoelting Company. https://doi.org/10.1037/t12018-000

Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *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. (2011). The creativity crisis: The decrease in creative thinking scores on the Torrance Tests of Creative Thinking. *Creativity Research Journal*, *23*, 285-295. https://doi.org/10.1080/10400419.2011.627805

Kirschenbaum, R. J. (1989). *Understanding the creative activity of students: Including an instruction manual for the creative behavior inventory*. Creative Learning Press.

Kirschenbaum, R. J. (1998). The creativity classification system: An assessment theory. *Roeper Review, 21*, 20-26. https://doi.org/10.1080/02783199809553921

Kiratli, N., Rozemeijer, F., Hilken, T., de Ruyter, K., & de Jong, A. (2016). Climate setting in sourcing teams: Developing a measurement scale for team creativity climate. *Journal of Purchasing and Supply Management, 22*, 196-204. https://doi.org/10.1016/j.pursup.2016.04.006

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1534-1555. https://doi.org/10.1037//0278-7393.25.6.1534

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

Kumar, V. K., & Holman, E. R. (1989). Creativity styles questionnaire. *Unpublished instrument*.

Lawshe, C. H., & Harris, D. H. (1960). *Manual of Instructions: Purdue Creativity Test*. Lafayette, Indiana, Purdue Research Foundation.

Lees-Haley, P. R. (1978). *Creative behavior inventory*. Huntsville, Alabama: Basic Research, Inc.

Long, H. (2014). An empirical review of research methodologies and methods in creativity studies (2003–2012). *Creativity Research Journal, 26*, 427-438.

Long, H. and J. Plucker (2015), "Assessing Creative Thinking: Practical applications". In Wegerif, R., L. Li and J. Kaufman (eds.), *The Routledge International Handbook of Research on Teaching Thinking*. Abingdon: Routledge. https://doi.org/10.1080/10400419.2014.961781

Long, H., Plucker, J. A., Yu, Q., Ding, Y., & Kaufman, J. C. (2014). Research productivity and performance of journals in the creativity sciences: A bibliometric analysis. *Creativity Research Journal*, *26*(3), 353-360. https://doi.org/10.1080/10400419.2014.929425

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*, 317-323. https://doi.org/10.1016/s0016-0032(26)91166-6

Lubart, T. I., Besançon, M., & Barbot, B. (2011). *Evaluation du potentiel créatif (EPoC)*. Paris, France: Editions Hogrefe.

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

Lucas, B., Claxton, G., & Spencer, E. (2013). *Progression in Student Creativity in School: First steps towards new forms of formative assessments* (OECD Education Working Paper, No. 86). Paris, France: OECD Publishing. https://doi.org/10.1787/19939019

Luo, J., & Niki, K. (2003). Function of hippocampus in “insight” of problem solving. *Hippocampus, 13*, 316-323. https://doi.org/10.1002/hipo.10069

MacGregor, J. N., & Cunningham, J. B. (2008). Rebus puzzles as insight problems. *Behavior Research Methods*, *40*, 263-268. https://doi.org/10.3758/brm.40.1.263

Madjar, N., & Ortiz-Walters, R. (2008). Customers as contributors and reliable evaluators of creativity in the service industry. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *29*(7), 949-966. https://doi.org/10.1002/job.522

Maltzman, I. (1960). On the training of originality. *Psychological Review*, *67*, 229-242. https://doi.org/10:1037h0046364

Martin, J., & Cox, D. (2016). Positioning Steve Nash: A Theory-Driven, Social Psychological, and Biographical Case Study of Creativity in Sport. *The Sport Psychologist, 30,* 388-398. https://doi.org/10.1123/tsp.2016-0002

Martindale, C., Hines, D., Mitchell, L., & Covello, E. (1984). EEG alpha asymmetry and creativity. *Personality and Individual Differences*, *5,* 77-86. https://doi.org/10.1016/0191-8869(84)90140-5

Mayer, R. E. (1995). Structural analysis of science prose: Can we increase problem-solving performance?. *ACM SIGDOC Asterisk Journal of Computer Documentation, 19*, 3-25. https://doi.org/10.1145/216708.216709

Mayer, R. E. (1999). Fifty years of creativity research. In R. J. Sternberg (Ed.), *Handbook of creativity*, 449 – 460. New York, NY: Cambridge University Press. https://doi.org/10.1017/cbo9780511807916.024

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

McDermid, C. D. (1965). Some correlates of creativity in engineering personnel. *Journal of Applied Psychology, 49*, 14-19. https://doi.org/10.1037/h0021658

Mednick, S. A., & Mednick, M. T. (1967). *Examiner's manual, Remote Associates Test: College and adult forms 1 and 2.* Houghton Mifflin.

Meeker, M. (1978). Measuring creativity from the child's point of view. *The Journal of Creative Behavior*, *12*, 52-62. https://doi.org/10.1002/j.2162-6057.1978.tb00159.x

Michael, W. B., & Colson, K. R. (1979). The development and validation of a life experience inventory for the identification of creative electrical engineers. *Educational and Psychological Measurement*, *39*, 463-470. https://doi.org/10.1177/001316447903900228

Miller, A. L. (2009). *Cognitive processes associated with creativity: Scale development and validation* (Doctoral dissertation, Ball State University).

Mumford, M. D., & Gustafson, S. B. (1988). Creativity syndrome: Integration, application, and innovation. Psychological Bulletin, *103*, 27–43. https://doi.org/10.1037//0033-2909.103.1.27

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

NACCCE, Great Britain. Dept. for Culture, Media, & Sport. (1999). *All our futures: Creativity, culture & education*. Dept. for Education and Employment.

Ng, T. W., & Feldman, D. C. (2012). A comparison of self-ratings and non-self-report measures of employee creativity. *Human Relations*, *65*(8), 1021-1047. https://doi.org/10.1177/0018726712446015

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

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

O'Neil, H. F., Abedi, J., & Spielberger, C. D. (1994). The measurement and teaching of creativity. *Motivation: Theory and research*, 245-263.

O'Quin, K., & Besemer, S. P. (1989). The development, reliability, and validity of the revised creative product semantic scale. *Creativity Research Journal, 2*, 267-278. https://doi.org/10.1080/10400418909534323

Owens, W. A. (1960). *The Owens' Creativity Test for Machine Design*. Iowa State University Press.

Owens, W. A., Schumacher, C. F., & Clark, J. B. (1957). The measurement of creativity in machine design. *Journal of Applied Psychology, 41*, 297-302. https://doi.org/10.1037/h0040668

Parkhurst, H. B. (1999). Confusion, lack of consensus, and the definition of creativity as a construct. *The Journal of Creative Behavior, 33*, 1-21. https://doi.org/10.1002/j.2162-6057.1999.tb01035.x

Plattner, H., Meinel, C., & Leifer, L. (Eds.). (2015). *Design thinking research: Making design thinking foundational.* Springer. https://doi.org/10.1007/978-3-642-21643-5

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational psychologist*, *39*, 83-96. https://doi.org/10.1207/s15326985ep3902_1

Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. *The Cambridge handbook of creativity*, 48-73. https://doi.org/10.1017/cbo9780511763205.005

Polya, G. (1945). *How to solve it: A new aspect of mathematical model.* Princeton, New Jersey: Princeton university press. https://doi.org/10.2307/2306109

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

Pringle, A., & Sowden, P. T. (2017). The Mode Shifting Index (MSI): A new measure of the creative thinking skill of shifting between associative and analytic thinking. *Thinking Skills and Creativity, 23,* 17-28. https://doi.org/10.1016/j.tsc.2016.10.010

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

Raudsepp, E. (1981). *How creative are you?*. Perigee.

Reid, J. B., King, F. J., & Wickwire, P. (1959). Cognitive and other personality characteristics of creative children. *Psychological Reports, 5,* 729-737. https://doi.org/10.2466/pr0.1959.5.3.729

Reisman, F., Keiser, L., & Otti, O. (2016). Development, use and implications of diagnostic creativity assessment app, RDCA–Reisman Diagnostic Creativity Assessment. *Creativity Research Journal*, *28*(2), 177-187. https://doi.org/10.1080/10400419.2016.1162643

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*(2), 144-152. https://doi.org/10.1037/aca0000227

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., & Westberg, K. L. (2002). *Scales for rating the behavioral characteristics of superior students. Technical and administration manual*. Creative Learning Press, Inc., PO Box 320, Mansfield, CT 06250.

Renzulli, J. S., & Reis, S. M. (1981). *Student product assessment form.* Mansfield Center, CT: Creative Learning Press. Retrieved from: http://teachers.renzullilearning.com/downloads/SPAF.pdf.

Rhoads, R. F., & Landy, F. J. (1973). Measurement of attitudes of industrial workgroups towards psychology and testing. *Journal of Applied Psychology, 58,* 197-201. https://doi.org/10.1037/h0035653

Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan*, *42*, 305-310.

Richards, R., Kinney, D. K., Benet, M., & Merzel, A. P. (1988). Assessing everyday creativity: Characteristics of the Lifetime Creativity Scales and validation with three large samples. *Journal of Personality and Social Psychology, 54,* 476-485. https://doi.org/10.1037//0022-3514.54.3.476

Richards Jr, J. M., Cline, V. B., & Needham, W. E. (1964). Creativity tests and teacher and self judgments of originality. *The Journal of Experimental Education, 32*, 281-285. https://doi.org/10.1080/00220973.1964.11010832

Richardson, C. P., & Saffle, M. (1983). Creativity research in music education: A review. *Bulletin of the Council for Research in Music Education*, 1-21.

Rimm, S. B. (1980). *Group inventory for finding creative talent (GIFT).* Watertown, WI: Educational Assessment Service.

Rimm, S. B. (1983). *PRIDE: Preschool and kindergarten interest descriptor*. Educational Assessment Service.

Robitzsch, A., & Steinfeld, J. (2018). Item response models for human ratings: Overview, estimation methods, and implementation in R. *Psychological Test and Assessment Modeling*, *60*(1), 101-138.

Roe, A. (1951). A psychological study of eminent biologists. *Psychological Monographs: general and applied*, *65*, i-68. ttps://doi.org/10.1037/h0093639

Rogaten, J., & Moneta, G. B. (2015). Development and validation of the short use of creative cognition scale in studying. *Educational Psychology, 35*(3), 294-314. https://doi.org/10.1080/01443410.2013.857011

Rookey, T. J. (1971). *The Pennsylvania assessment of creative tendency: Norms-technical manual*. Pennsylvania: Pennsylvania Department of Education.

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. (1989). Parents' and teachers' ratings of the creativity of children. *Journal of Social Behavior and Personality*, *4*, 73-84.

Runco, M. (2004). Personal creativity and culture. In *Creativity: When east meets west* (pp. 9-21). https://doi.org/10.1142/9789812567192_0002

Runco, M. A., Abdulla, A. M., Paek, S. H., Al-Jasim, F. A., & Alsuwaidi, H. N. (2016). Which test of divergent thinking is best*?. Creativity. Theories–Research-Applications*, *3*, 4-18. https://doi.org/10.1515/ctra-2016-0001

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

Runco, M. A., Millar, G., Acar, S., & Cramond, B. (2010). Torrance tests of creative thinking as predictors of personal and public achievement: A fifty-year follow-up. *Creativity Research Journal*, *22*, 361-368. https://doi.org/10.1080/10400419.2010.523393

Runco, M. A., Plucker, J. A., & Lim, W. (2001). Development and psychometric integrity of a measure of ideational behavior. *Creativity Research Journal*, *13*, 393-400. https://doi.org/10.1207/s15326934crj1334_16

Runco Creativity Assessment Battery (rCAB). (2011). Creativity Testing Services (CTS). Retrieved fromhttp://creativitytestingservices.com/

Rusch, R. R., Denny, D. A., & Ives, S. (1964). The development of a test of creativity in the dramatic arts: a pilot study. *The Journal of Educational Research, 57*, 250-254. https://doi.org/10.1080/00220671.1964.10883073

Said-Metwaly, S., Van den Noortgate, W., & Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. *Creativity. Theories–Research-Applications, 4*(2), 238-275. https://doi.org/10.1515/ctra-2017-0013

Schack, G. D. (1989). Self-efficacy as a mediator in the creative productivity of gifted children. *Journal for the Education of the Gifted*, *12*, 231-249. https://doi.org/10.1177/016235328901200306

Schaefer, C. E., & Anastasi, A. (1968). A biographical inventory for identifying creativity in adolescent boys. *Journal of Applied Psychology, 52*, 42-48. https://doi.org/10.1037/h0025328

Schaefer, C. E., & Bridges, C. I. (1970). Development of a creativity attitude survey for children. *Perceptual and Motor Skills, 31*, 861-862. https://doi.org/10.2466/pms.1970.31.3.861

Schmeidler, G. R. (1965). Visual imagery correlated to a measure of creativity. *Journal of Consulting Psychology*, *29*, 78-80. https://doi.org/10.1037/h0021660

Schultz, C. B., & Skager, R. W. (1963). Relationship of an independent activities questionnaire to performance during high school. *ETS Research Report Series*, *1963*(1), i-19. https://doi.org/10.1002/j.2333-8504.1963.tb00310.x

Shaw, G. A., & Conway, M. (1990). Individual differences in nonconscious processing: The role of creativity. *Personality and Individual Differences, 11*, 407-418. https://doi.org/10.1016/0191-8869(90)90224-f

Sheth, B. R., Sandkühler, S., & Bhattacharya, J. (2009). Posterior beta and anterior gamma oscillations predict cognitive insight. *Journal of Cognitive Neuroscience, 21*, 1269-1279. https://doi.org/10.1162/jocn.2009.21069

Siegel, L. (1956). A biographical inventory for students: I. Construction and standardization of the instrument. *Journal of Applied Psychology*, *40*, 5-10. https://doi.org/10.1037/h0040589

Silvia, P. J., Wigert, B., Reiter-Palmon, R., & Kaufman, J. C. (2012). Assessing creativity with self-report scales: A review and empirical evaluation. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(1), 19-34. https://doi.org/10.1037/a0024071

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*(5), 328-340. https://doi.org/10.1016/j.intell.2013.05.004

Simonton, D. K. (1977). Eminence, creativity, and geographical marginality: A recursive structural equation model. *Journal of Personality and Social Psychology, 35*, 805-816. https://doi.org/10.1037//0022-3514.35.11.805

Simonton, D. K. (1998). Fickle fashion versus immortal fame: Transhistorical assessments of creative products in the opera house. *Journal of Personality and Social Psychology*, *75*, 198-210. https://doi.org/10.1037//0022-3514.75.1.198

Skager, R. W., Schultz, C. B., & Klein, S. P. (1965). Quality and quantity of accomplishments as measures of creativity. *Journal of Educational Psychology, 56*, 31-39. https://doi.org/10.1037/h0021901

Slahova, A., Savvina, J., Cacka, M., & Volonte, I. (2007). Creative activity in conception of sustainable development education. *International Journal of Sustainability in Higher Education*, *8*(2), 142-154. https://doi.org/10.1108/14676370710726616

Snyder, H. T., Hammond, J. A., Grohman, M. G., & Katz-Buonincontro, J. (2019). Creativity measurement in undergraduate students from 1984–2013: A systematic review. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 133-143. https://doi.org/10.1037/aca0000228

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

Stankov, L. (1999). Mining on the "no man's land" between intelligence and personality. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 315-337). Washington, DC, US: American Psychological Association. http://dx.doi.org/10.1037/10315-014

Starkweather, E. K. (1971). Creativity research instruments designed for use with preschool children. *The Journal of Creative Behavior, 5*, 245-255. https://doi.org/10.1002/j.2162-6057.1971.tb00895.x

Starkweather, E. K. (1974). Starkweather Originality Test for Young Children.

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

Sternberg, R. J. (2009). The Rainbow and Kaleidoscope Projects: A new psychological approach to undergraduate admissions. *European Psychologist*, *14*, 279-287. https://doi.org/10.1027/1016-9040.14.4.279

Sternberg, R. J. (2012). The assessment of creativity: An investment-based approach. *Creativity research journal*, *24*, 3-12. https://doi.org/10.1080/10400419.2012.652925

Stricker, L. J., Rock, D. A., & Bennett, R. E. (2001). Sex and ethnic-group differences on accomplishments measures. *Applied Measurement in Education, 14,* 205-218. https://doi.org/10.1207/s15324818ame1403_1

Tagiuri, R. (1960). Movement as a cue in person perception. In *Perspectives in personality research* (p. 175-195). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-39598-1_9

Taylor, C. W., Smith, W. R., & Ghiselin, B. (1963). The creative and other contributions of one sample of research scientists. In *Scientific creativity: its recognition and development.* (p. 63-81). New York: Wiley.

Taylor, I. A., & Fish, T. A. (1979). The Creative Behaviour Disposition Scale: A Canadian validation. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 11*, 95-97. https://doi.org/10.1037/h0081577

Taylor, I. A., & Sandler, B. E. (1972). Use of a creative product inventory for evaluating products of chemists. *Proceedings of the Annual Convention of the American Psychological Association, 7(Pt. 1), 311-312*.

Taylor, I. A. (1975). An emerging view of creative actions. In I. A. Taylor & J. W. Getzels (Eds.), *Perspectives in creativity* (pp. 297–325). Chicago: Aldine.

Taylor, I. A., Sutton, D., & Haworth, S. (1974). The measurement of creative transactualization: A scale to measure behavioral dispositions to creativity. *Journal of Creative Behavior*, *8*, 114-115. https://doi.org/10.1002/j.2162-6057.1974.tb01118.x

Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements.* New York: Teachers College, Columbia University.

Tierney, P., & Farmer, S. M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management Journal, 45*, 1137-1148. https://doi.org/10.2307/3069429

Tierney, P., Farmer, S. M., & Graen, G. B. (1999). An examination of leadership and employee creativity: The relevance of traits and relationships. *Personnel Psychology, 52*, 591-620. https://doi.org/10.1111/j.1744-6570.1999.tb00173.x

Torrance, E. P. (1975). Assessing children, teachers, and parents against the ideal child criterion. *Gifted Child Quarterly, 19*, 130-139. https://doi.org/10.1177/001698627501900220

Torrance, E. P. (1969). Creativity. What Research Says to the Teacher, Series, No. 28.

Torrance, E. P. (1962). Guiding creative talent. https://doi.org/10.1037/13134-000

Torrance, E. P. (1969). Prediction of adult creative achievement among high school seniors. *Gifted Child Quarterly, 13*, 223-229. https://doi.org/10.1177/001698626901300401

Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking—Norms, Technical Manual Research Edition—Verbal Tests, Forms A and B—Figural Tests, Forms A and B.* Princeton, NJ: Personnel Press.

Torrance, E. P., & Gibbs, S. (1977). *Administration, Scoring, and Norms Manual, Thinking Creatively in Action and Movement*. Department of Educational Psychology, University of Georgia.

Torrance, E. P., & Khatena, J. (1970). "What Kind of Person Are You?" A Brief Screening Device for Identifying Creatively Gifted Adolescents and Adults. *Gifted Child Quarterly, 14*, 71-75. https://doi.org/10.1177/001698627001400201

Torrance, E. P., Khatena, J., & Cunnington, B. F. (1973). *Thinking Creatively with Sounds and Words: Sounds and Images; Onomatopoeia and Images*. Personnel Press/Ginn.

Vallée-Tourangeau, F. (Ed.). (2018). *Insight: on the origins of new ideas*. Routledge. https://doi.org/10.4324/9781315268118

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

Vincent-Lancrin, S. (2017, May). *Future learning skills in a global context: from discourse to practice*. http://docplayer.net/51475061-Future-learning-skills-in-a-global-context-from-discourse-to-practice.html.

Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction.* Oxford, England: Holt, Rinehart & Winston.

Wallen, N. E., & Stevenson, G. M. (1960). Stability and correlates of judged creativity in fifth grade writings. *Journal of Educational Psychology, 51*, 273-276. https://doi.org/10.1037/h0045933

Ward, W. C., & Cox, P. W. (1974). A field study of nonverbal creativity. *Journal of personality*, *42*, 202-219. https://doi.org/10.1111/j.1467-6494.1974.tb00670.x

Webster, P. R. (1994). *Measure of Creative Thinking in Music:(MCTM): Administrative Guidelines*. Peter Webster.

Wechsler, S. M., Vendramini, C. M. M., & Oakland, T. (2012). Thinking and creative styles: A validity study. *Creativity Research Journal*, *24*, 235-242. https://doi.org/10.1080/10400419.2012.677359

Weisberg, R. W. (1988). Problem solving and creativity. *The nature of creativity,* 148-176.

Weiss, S., & Roberts, R. D. (2018). Using anchoring vignettes to adjust self-reported personality: A comparison between countries. *Frontiers in Psychology*, *9*, 325. https://doi.org/10.3389/fpsyg.2018.00325

Weiss, S., Steger, D., Kaur, Y., Hildebrandt, A., Schroeders, U., & Wilhelm, O. (2020). On the trail of creativity: Dimensionality of divergent thinking and its relation with cognitive abilities and personality. *European Journal of Personality, 0*(0), 1-22. https://doi.org/10.1002/per.2288

Weiss, S., & Wilhelm, O. (2020). Coda: Creativity in psychological research versus in linguistics – Same but different? *Cognitive Semiotics*, *13*(1), 1-15. https://doi.org/10.1515/cogsem-2020-2029

Welsh, G. S., & Barron, F. (1959). *Barron-Welsh art scale*. Consulting Psychologists Press.

Westby, E. L., & Dawson, V. L. (1995). Creativity: Asset or burden in the classroom?. *Creativity Research Journal*, *8*, 1-10. https://doi.org/10.1207/s15326934crj0801_1

White, A., Shen, F., & Smith, B. L. (2002). Judging advertising creativity using the creative product semantic scale. *The Journal of Creative Behavior, 36*, 241-253. https://doi.org/10.1002/j.2162-6057.2002.tb01067.x

Williams, F. E. (1993). *Creativity Assessment Packet: CAP*. Pro-Ed.

Wilson, R. C., Guilford, J. P., Christensen, P. R., & Lewis, D. J. (1954). A factor-analytic study of creative-thinking abilities. *Psychometrika*, *19*, 297-311. https://doi.org/10.1007/bf02289230

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

Woo, S. E., Chernyshenko, O. S., Stark, S. E., & Conz, G. (2014). Validity of six openness facets in predicting work behaviors: A meta-analysis. *Journal of Personality Assessment, 96*, 76-86. https://doi.org/10.1080/00223891.2013.806329

Yamamoto, K. (1964). *Experimental Scoring Manuals for Minnesota Tests of Creative Thinking and Writing*. Kent, OH: Bureau of Educational Research, Kent State University.

Zhang, X., & Bartol, K. M. (2010). Linking empowering leadership and employee creativity: The influence of psychological empowerment, intrinsic motivation, and creative process engagement. *Academy of Management Journal*, *53*(1), 107-128. https://doi.org/10.5465/amj.2010.48037118

Zhou, J., & George, J. M. (2001). When job dissatisfaction leads to creativity: Encouraging the expression of voice. *Academy of Management Journal*, *44*(4), 682-696. https://doi.org/10.5465/30694

IV. Manuscript III

On the Trail of Creativity: Dimensionality of Divergent Thinking and its Relation with Cognitive Abilities and Personality

The experimental idea and study design were developed by all co-authors, mainly Prof. Dr. Andrea Hildebrandt, Prof. Dr. Oliver Wilhelm, and Prof. Dr. Ulrich Schroeders. All co-authors contributed to the programming. The organization and data collection was conducted by myself, Dr. Diana Steger, and Yadwinder Kaur. The data analysis and discussion of results were conducted by myself under the supervision of Prof. Dr. Oliver Wilhelm, Prof. Dr. Andrea Hildebrandt, and Prof. Dr. Ulrich Schroeders. I have done a literature search and drafted the manuscript. All co-authors have revised the manuscript. The manuscript was accepted for publication in the European Journal of Personality on the 22nd of September 2020. The article is published open access, including the following rights: Open Access CC BY 4.0, https://creativecommons.org/licenses/by/4.0/

Weiss, S., Steger, D., Kaur, Y., Hildebrandt, A., Schroeders, U., & Wilhelm, O. (2020). On the trail of creativity: Dimensionality of divergent thinking and its relation with cognitive abilities and personality. *European Journal of Personality*. https://doi.org/10.1002/per.2288

European Journal of Personality, *Eur. J. Pers.* (2020)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/per.2288

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]

[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*

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. © 2020 The Authors. European Journal of Personality published by John Wiley & Sons Ltd on behalf of European Association of Personality Psychology

Key words: divergent thinking; intelligence; personality; insight

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, 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).

*Correspondence to: Selina Weiss, Institute for Psychology and Pedagogy, Ulm University, Albert-Einstein Strasse 47, 89081 Ulm, Germany.
E-mail: selina.weiss@uni-ulm.de
This research was supported by the German Research Foundation (HI1780/2–1) with grants awarded to Andrea Hildebrandt and supported by the Bamberg Graduate School of Social Sciences, which is funded by the German Research Foundation (DFG) under the German Excellence Initiative (GSC1024). We thank the numerous research assistants who helped during the data acquisition and human coding.For all studies, we have reported conditions and data exclusions. Decisions surrounding final sample sizes are presented in the sample sections. Research objectives and hypothesis were not preregistered. The data are available along with syntax for the statistical analysis at https://osf.io/c8j29/.
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/. Author's disclosure form may also be found at the Supporting Information in the online version.

Handling editor: Christian Kandler
Received 26 August 2019
Revised 22 June 2020, Accepted 22 June 2020

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).

[1]According to a Darwinian theory (Campbell, 1960; Simonton, 1999), the creative process includes two mental mechanisms: *blind variation* and *selective retention*. More recent work argues that the theoretical link between these Darwinian mental processes and creativity is problematic (Gabora, 2011).

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

DOI: 10.1002/per

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, Mühlmann, 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

DOI: 10.1002/per

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 (β = .29) and common executive functions like inhibition (β = .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

DOI: 10.1002/per

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 (β = −.04) and conscientiousness and creativity (β = −.04; Silvia, Kaufman, Reiter-Palmon, & Wigert, 2011). It did however uncover a small negative association between honesty–humility and creativity (β = −.20), a small relation between creativity and extraversion (β = .17), and a strong association between openness and creativity (β = .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

DOI: 10.1002/per

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.

[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 in *Intelligence*.

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

DOI: 10.1002/per

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 interrater 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 (WMv) and figural (WMf) 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*

DOI: 10.1002/per

(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

DOI: 10.1002/per

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		−.02	.00	**−.19**	−.08	.17	.13	−.15	−.04	**−.21**	−.17	−.05	**−.20**	**.27**	.09	**.31**
		(3.78)		(.83)	(.98)	**(.04)**	(.41)	(.07)	(.17)	(.11)	(.66)	**(.02)**	(.07)	(.57)	**(.03)**	**(.00)**	(.35)	**(.00)**
2	Sex	1.46		**−.24**	**−.19**	**−.28**	**−.22**	.11	−.17	−.18	−.09	**−.26**	**−.23**	−.02	−.13	**.23**	.11	**.25**
		(.50)		**(.01)**	**(.04)**	**(.00)**	**(.02)**	(.22)	(.07)	(.06)	(.35)	**(.01)**	**(.01)**	(.79)	(.16)	**(.01)**	(.26)	**(.01)**
3	SA	6.79	.84		**.39**	**.54**	**.48**	**.39**	**.27**	.06	.00	.04	−.02	**.20**	.18	**.23**	**.22**	.14
		(2.15)			**(.00)**	**(.00)**	**(.00)**	**(.00)**	**(.00)**	(.53)	(.98)	(.66)	(.87)	**(.03)**	(.06)	**(.01)**	**(.02)**	(.13)
4	IN	.77	.90			**.29**	**.28**	**.24**	**.30**	.06	−.04	.08	.07	.13	.15	**.19**	**.23**	.12
		(.38)				**(.00)**	**(.00)**	**(.01)**	**(.00)**	(.54)	(.71)	(.38)	(.45)	(.15)	(.12)	**(.04)**	**(.01)**	(.18)
5	FF	5.24	.70				**.35**	.05	**.19**	.17	.05	**.24**	.13	**.27**	**.22**	.02	.00	−.06
		(1.43)					**(.00)**	(.56)	**(.04)**	(.07)	(.58)	**(.01)**	(.15)	**(.00)**	**(.02)**	(.85)	(.98)	(.51)
6	RF	8.29	.84					**.30**	**.26**	**.20**	**.20**	.15	.17	.11	**.35**	.15	**.21**	.05
		(2.45)						**(.00)**	**(.00)**	**(.03)**	**(.03)**	(.12)	(.07)	(.25)	**(.00)**	(12.)	**(.03)**	(.56)
7	CO	1.80	.64						.18	−.10	−.11	−.08	−.11	.08	.17	**.38**	**.27**	.18
		(.40)							(.06)	(.30)	(.25)	(.42)	(.22)	(.41)	(.06)	**(.00)**	**(.00)**	(.05)
8	NI	2.04	.75							.10	.07	.00	.09	.16	**.21**	.12	**.26**	**.20**
		(.66)								(.27)	(.44)	(.99)	(.33)	(.08)	**(.03)**	(.21)	**(.01)**	**(.03)**
9	WM_f	.74									**.27**	.18	.19	**.31**	**.38**	−.06	−.03	−.02
		(.14)									**(.00)**	(.06)	(.04)	**(.00)**	**(.00)**	(.54)	(.73)	(.83)
10	WM_v	.77										.13	.13	**.27**	**.33**	−.02	.03	.06
		(.14)										(.15)	(.15)	**(.00)**	**(.00)**	(.83)	(.73)	(.49)
11	MS_f	.91											**.79**	.11	.16	−.15	−.08	−.17
		(.16)											**(.00)**	(.24)	(.08)	(.11)	(38)	(.07)
12	MS_v	1.08												.03	**.22**	−.09	−.07	−.08
		(.19)												(.75)	**(.02)**	(.33)	(.48)	(.38)
13	gf_f	.54	.67												**.39**	**.23**	.13	**.20**
		(.17)													**(.00)**	**(.01)**	(.17)	**(.03)**
14	gf_v	.68	.63													**.26**	.11	.14
		(.16)														**(.01)**	(.25)	(.12)
15	gc_{social}	.68															**.39**	**.41**
		(.18)															**(.00)**	**(.00)**
16	gc_{human}	.81																**.32**
		(.15)																**(.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. $\omega_{WM} = .47$, $\omega_{MS} = .83$, and $\omega_{gc} = .65$

DOI: 10.1002/per

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 (.01)**
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)**	**.15 (.02)**
6	RF					**.33 (.00)**	**.16 (.01)**	**.17 (.01)**	**.39 (.00)**	**.26 (.00)**	.07 (.33)	**.20 (.00)**	.12 (.05)	**.15 (.03)**	**.17 (.01)**	**.20 (.00)**
7	CO						**.33 (.00)**	.05 (.35)	**.15 (.01)**	**.17 (.00)**	.02 (.62)	**.18 (.00)**	**.15 (.02)**	.09 (.14)	**.23 (.00)**	.06 (.27)
8	NI							**.13 (.03)**	.11 (.09)	**.26 (.00)**	.01 (.93)	.06 (.35)	.02 (.74)	.09 (.16)	.08 (.22)	.03 (.66)
9	SCRorg								**.52 (.00)**	**.33 (.00)**	.12 (.07)	**.15 (.02)**	−.04 (.55)	.00 (.98)	**.19 (.00)**	.06 (.34)
10	SCRflu									**.46 (.00)**	**.20 (.00)**	**.23 (.00)**	.04 (.59)	.04 (.47)	**.25 (.00)**	.12 (.06)
11	ANAorg										**.19 (.00)**	**.29 (.00)**	.09 (.15)	.10 (.14)	**.27 (.00)**	**.20 (.00)**
12	WM_f											**.35 (.00)**	**.13 (.04)**	−.04 (.54)	**.30 (.00)**	**.23 (.00)**
13	WM_v												**.22 (.00)**	**.19 (.00)**	**.48 (.00)**	**.37 (.00)**
14	MS_f													**.71 (.00)**	**.22 (.00)**	**.20 (.00)**
15	MS_v														**.17 (.00)**	**.20 (.00)**
16	gf_f															**.43 (.00)**
17	gf_v															

DOI: 10.1002/per

Table 2. (Continued)

		Mean (*SD*)	ω	18	19	20	21	22	23	24	25	26	27	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)**	−.12 (.07)	−.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)**	.05 (.35)	−.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)	.05 (.44)	.04 (.49)
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)	**.17 (.01)**	−.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)	.12 (.05)	.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)	−.02 (.78)	.02(.73)	.00 (.98)	.00 (.90)	.06 (.30)
18	gc_{nature}	.58 (.15)			**.33 (.00)**	**.32 (.00)**	**.38 (.00)**	.12 (.05)	**.20 (.00)**	−.03 (.60)	.06 (.38)	−.08 (.21)	.11 (.07)	.05 (.42)
19	gc_{life}	.60 (.13)				**.42 (.00)**	**.38 (.00)**	**.22 (.00)**	.11 (.08)	.02 (.76)	.03 (.66)	−.05 (.40)	−.01 (.92)	.04 (.54)
20	gc_{human}	.58 (.13)					**.51 (.00)**	**.35 (.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); MS, mental speed (f, figural; v, verbal); WM, working memory capacity (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 $\omega_{gc} = .72$

DOI: 10.1002/per

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 ≥ .95, RMSEA ≤ .06, and SRMR ≤ .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 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)

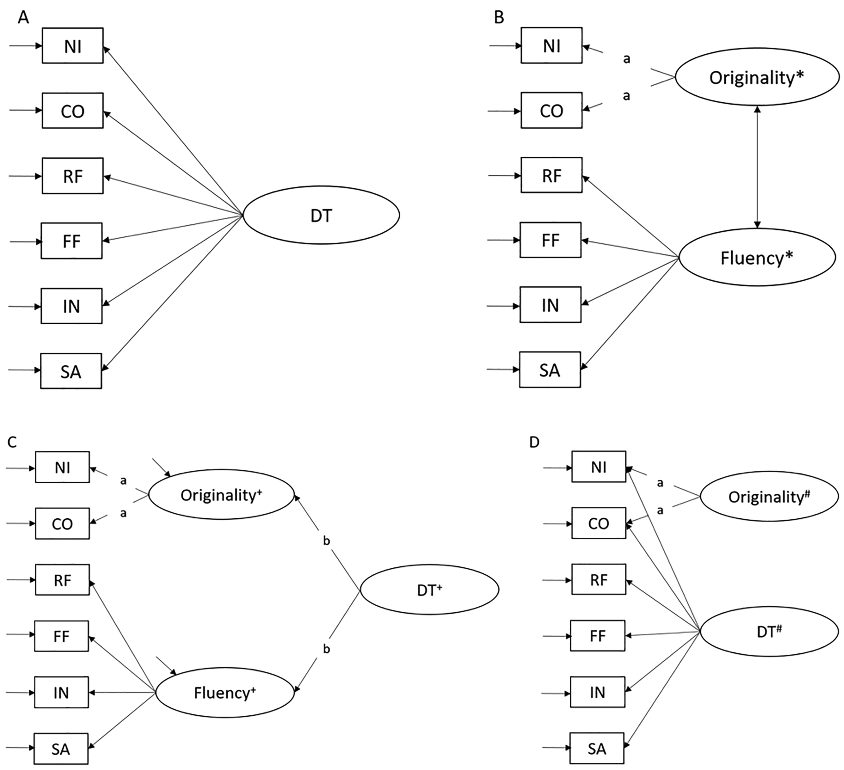


Figure 1. (A–D) Competing measurement models of divergent thinking (DT). Indicators are based on test scores computed as described in the method section. Fluency indicators are SA (similar attributes), IN (inventing names), FF (figural fluency), and RF (retrieval fluency). CO (combining objects) and 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).

DOI: 10.1002/per

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 (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 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 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

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

Model	χ^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 DT^{+} = .65
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.

DOI: 10.1002/per

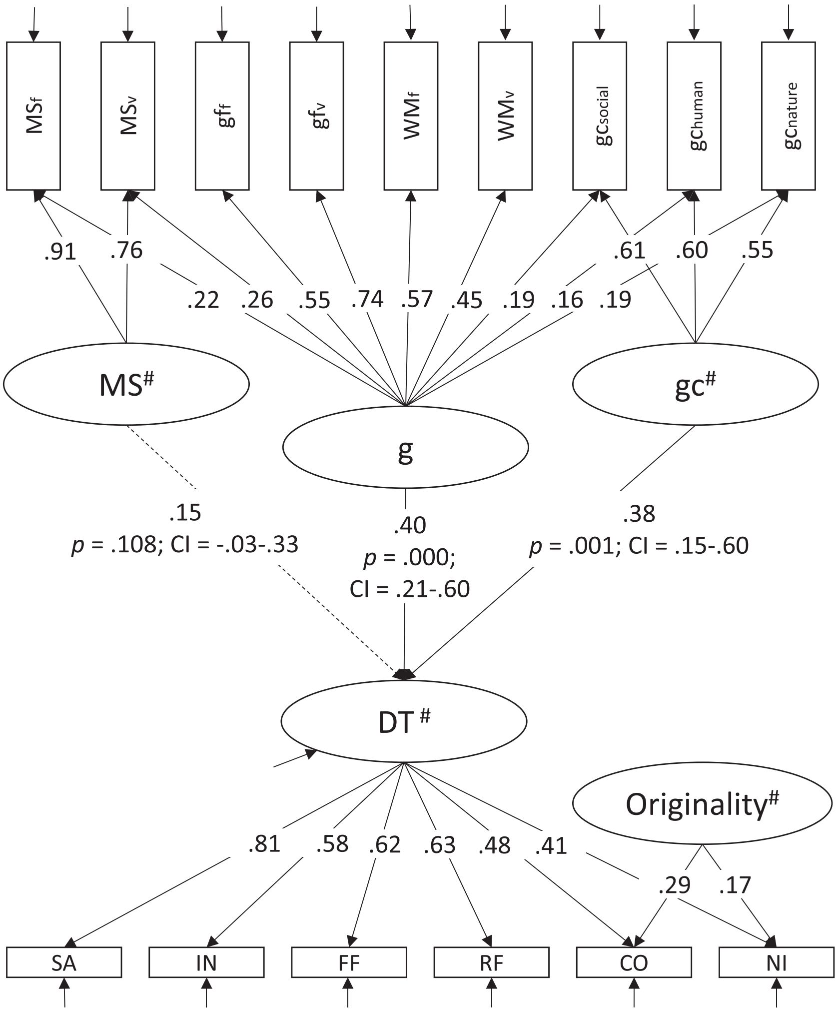


Figure 2. Structural model (Study 1; $N = 152$) relating DT# to general cognitive ability (g), crystallized intelligence (gc#), and mental speed (MS#). Nonsignificant latent regressions are displayed as dotted lines. All coefficients are standardized. Nested factors in the cognitive ability model are 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.

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

Model	χ^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.

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

DOI: 10.1002/per

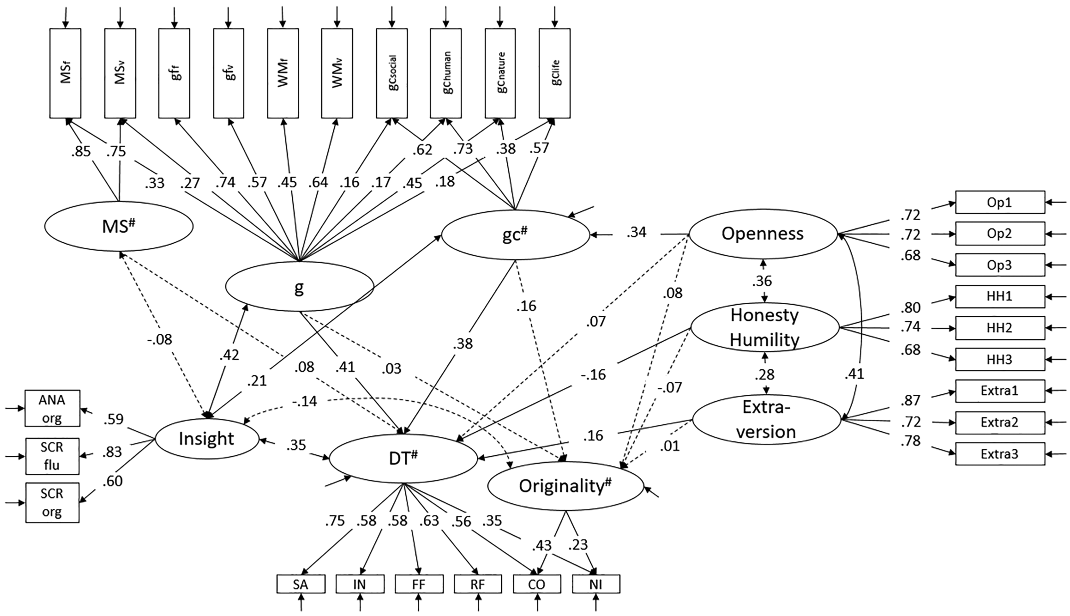


Figure 3. Structural model (Study 2; N = 298) relating DT# and originality# to general cognitive ability (g), gc#, and MS# insight and personality traits. Nonsignificant latent regressions are displayed as dotted lines. All coefficients are standardized. Nested factors in the cognitive ability model 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.

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 (ω = .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).

DOI: 10.1002/per

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

DOI: 10.1002/per

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, Bürkner, & 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; Käckenmester, 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

DOI: 10.1002/per

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.

ACKNOWLEDGEMENT

Open access funding enabled and organized by Projekt DEAL.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Descriptive Statistics for all Creativity Indicators Including Inter-Rater Reliabilities (ICCs) for Study 1
Table S2. Descriptive Statistics for all Creativity Indicators Including Inter-Rater Reliabilities (ICCs) for Study 2
Table S3. Descriptive Statistics for all Insight Indicators Including Inter-Rater Reliabilities (ICCs) for Study 2
Table S4. Fit Indices of the Measurement Models of all Creativity Tasks on the Item Level for Study 1
Table S5. Fit Indices of the Measurement Models of all Creativity Tasks on the Item Level for Study 2
Table S6. Correlations and Regressions between Latent Variables for the Model Depicted in Figure 3 When all Links Are Allowed Except Relations between Superordinate and Associated Subordinate Factors
Table S7. Correlations and Regressions between Latent Variables for the Model Depicted in Figure 3
Figure S1A. Correlations and bivariate scatterplots for Study 1 between the manifest scores for fluency and originality, fluid intelligence (gff = figural, gfv = verbal) and crystallized intelligence (gc); * < .05; ** < .01; *** < .001.
Figure S1B. Correlations and bivariate scatterplots for Study 2 between the manifest scores for fluency and originality, fluid intelligence (gff = figural, gfv = verbal) and crystallized intelligence (gc), insight, and personality (openness, honesty-humility [HonHum] and extraversion); * < .05; ** < .01; *** < .001.
Figure S2. Competing measurement models of DT as depicted in Figure 1 in the paper including standardized loadings for Study1/Study2. Indicators are 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. The factor variances of the latent variables were fixed to 1. All factors were scaled using unit variance identification constraints (Kline, 2015).

DOI: 10.1002/per

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.actclub.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 HEXACO-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/10.1016/j.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/MONO.132.4.355-430

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., Bürkner, 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., Mühlmann, 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/CBO9780511571312

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. Vallée-Tourangeau (Ed.), *Insight: On the origins of new ideas* (1st ed., pp. 79–104). Routledge https://doi.org/10.4324/9781315268118-5.

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/10.1016/j.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/jintelligence6030042
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., Günther, 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/h0023296
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/10.1037/0022-0663.71.2.191
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/10.1080/10705519909540118 7
Jäger, A. O. (2006). *Berliner intelligenzstruktur-test für 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/CBO9780511977244.039
Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *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)—Worträtseln 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/10.1037/0033-2909.114.1.185

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 persönlichkeitsinventars. *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/10.1016/j.paid.2005.06.022

Preckel, F., Wermer, C., & Spinath, F. M. (2011). The interrelationship 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/10.1037/0882-7974.19.1.40

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 kreativitäts-test-VKT: Ein verfahren zur crfassung verbal-produktiver kreativitätsmerkmale [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/10.1177/1073191120914970 107319112091497.

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 of creativity*. 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 für 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

V. Manuscript IV

A Reappraisal of the Threshold Hypothesis of Creativity and Intelligence

I and Diana Steger developed the conceptualization of the paper under the supervision of Prof. Dr. Ulrich Schroeders and Prof. Dr. Oliver Wilhelm. The data analysis discussion of results was conducted by myself under the supervision of Prof. Dr. Oliver Wilhelm, Prof. Dr. Ulrich Schroeders, and Diana Steger. The literature research was done by myself. I drafted the original manuscript, and all co-authors revised the manuscript. The manuscript was accepted for publication in the Journal of Intelligence on 3rd of November 2020. The article is published open access, including the following rights:

Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020). A reappraisal of the threshold hypothesis of creativity and intelligence. *Journal of Intelligence*, *8*(4), 38-58. https://doi.org/10.3390/jintelligence8040038

Journal of *Intelligence*

MDPI

Article

A Reappraisal of the Threshold Hypothesis of Creativity and Intelligence

Selina Weiss [1,*], Diana Steger [1], Ulrich Schroeders [2] and Oliver Wilhelm [1]

[1] Institute of Psychology and Pedagogy, Ulm University, Albert-Einstein Allee 47, 89081 Ulm, Germany; diana.steger@uni-ulm.de (D.S.); oliver.wilhelm@uni-ulm.de (O.W.)

[2] Institute of Psychology, University of Kassel, Holländische Strasse 36-38, 34127 Kassel, Germany; schroeders@psychologie.uni-kassel.de

* Correspondence: selina.weiss@uni-ulm.de

Received: 28 August 2020; Accepted: 3 November 2020; Published: 11 November 2020

check for updates

Abstract: Intelligence has been declared as a necessary but not sufficient condition for creativity, which was subsequently (erroneously) translated into the so-called threshold hypothesis. This hypothesis predicts a change in the correlation between creativity and intelligence at around 1.33 standard deviations above the population mean. A closer inspection of previous inconclusive results suggests that the heterogeneity is mostly due to the use of suboptimal data analytical procedures. Herein, we applied and compared three methods that allowed us to handle intelligence as a continuous variable. In more detail, we examined the threshold of the creativity-intelligence relation with (a) scatterplots and heteroscedasticity analysis, (b) segmented regression analysis, and (c) local structural equation models in two multivariate studies (N_1 = 456; N_2 = 438). We found no evidence for the threshold hypothesis of creativity across different analytical procedures in both studies. Given the problematic history of the threshold hypothesis and its unequivocal rejection with appropriate multivariate methods, we recommend the total abandonment of the threshold.

Keywords: creativity; intelligence; threshold hypothesis; necessary but not sufficient condition

1. Introduction

Can you be creative without being smart? Many researchers argued that creativity presupposes intelligence (e.g., Guilford 1967) and intuitively this proposition probably makes sense for many readers. Indeed, the abilities needed for divergent production/thinking (Guilford 1967) and idea generation and evaluation (Mumford and McIntosh 2017) are closely intertwined with other cognitive abilities, commonly referred to as convergent thinking (Carroll 1993; Cropley 2006). For example, the creativity required to come up with an invention for high-tech problems builds upon substantial expertise in a field, as well as decontextualized fluid intelligence (e.g., Nusbaum and Silvia 2011). However, the intellectual prerequisites for different tasks challenging creativity might vary (Diedrich et al. 2018; Jauk et al. 2013) and the relevance of general intelligence might not be the same at different points in the distribution of creative abilities.

Historically, creative ability was incorporated in most models of intelligence, predominantly as a lower-order factor below general intelligence. Creative abilities[1]—often measured by divergent thinking tasks, including indicators of fluency or originality (Runco 2008)—are part of the structure of intellect model (Guilford 1967), the three-stratum theory of cognitive abilities (Carroll 1993), and the

[1] We understand creativity as the ability to produce divergent ideas; thus, we do not further distinguish between creativity and divergent thinking for the purpose of this paper and use the terms interchangeably from now on.

Berlin intelligence structure model (Jäger et al. 1997; Süß and Beauducel 2005). The relation between intelligence and creativity was evaluated in several studies (e.g., in terms of a lower-order factor in the Cattell–Horn–Carroll model of cognitive abilities; McGrew 2009; Silvia et al. 2013). Recent evidence showed that creative abilities (e.g., divergent thinking scored for fluency) and general intelligence were substantially related ($r = 0.46$, Karwowski et al. 2018; $\beta = 0.45$, Nusbaum and Silvia 2011), especially when using a broad set of indicators ($\beta = 0.51$; Benedek et al. 2012; $\beta = 0.40$, Weiss et al. 2020a). This is corroborated by a review that states that the progress in analytical tools, as well as in measurement (e.g., in cognitive neuroscience), has led to the conclusion that creativity and intelligence are closely related (Silvia 2015). Research reporting lower correlations are often based either on narrow measures of the constructs or on very heterogenous measures (e.g., a meta-analysis by Kim (2005) found a mean correlation of $r = 0.17$). Among others, the substantial correlation between the two constructs resurrected the question if the relation between creativity and intelligence might not follow a necessary condition, but a necessary but not sufficient condition (Guilford 1967). Further, they wondered if it was in accordance with the so-called threshold hypothesis (e.g., Karwowski et al. 2016). In the present paper, we reviewed different interpretations of Guilford's original finding and tried to translate them to testable statistical means. Moreover, we discussed three analytical approaches to study the relation between intelligence and creativity in two different data sets that varied with regard to the age of samples and the measures for creativity and intelligence.

2. The Threshold Hypothesis of Creativity and Intelligence

Guilford was one of the first to describe and investigate the relationship between creativity and intelligence. In his initial publication, he stated that "high IQ is not a sufficient condition for high DP [divergent production] ability; it is almost a necessary condition" (Guilford 1967, p. 168). Thus, Guilford assumed that highly intelligent individuals are not necessarily creative but can be creative, while less intelligent individuals are necessarily less creative (Guilford and Christensen 1973), which became an assumption known as the necessary but not sufficient condition. This relationship is schematically depicted in the left plot in Figure 1. If Guilford's assumption holds and intelligence is a necessary but not sufficient condition for being creative, individuals' scores scatter within the triangle. Although the original wording of Guilford's theory was quite unambiguous, comparatively little research was done to test this assumption. Only recently, researchers picked up on the necessary but not sufficient condition (e.g., Karwowski et al. 2016; Shi et al. 2017). In contrast to the necessary but not sufficient condition, one can see that the necessary (and sufficient) condition corresponds to an ordinary linear regression (see Figure 1, middle plot).

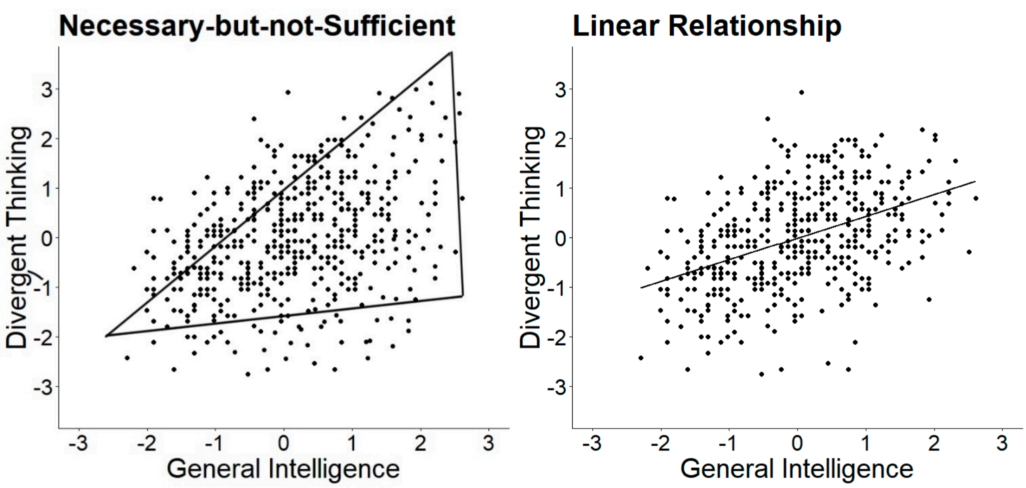


Figure 1. *Cont.*

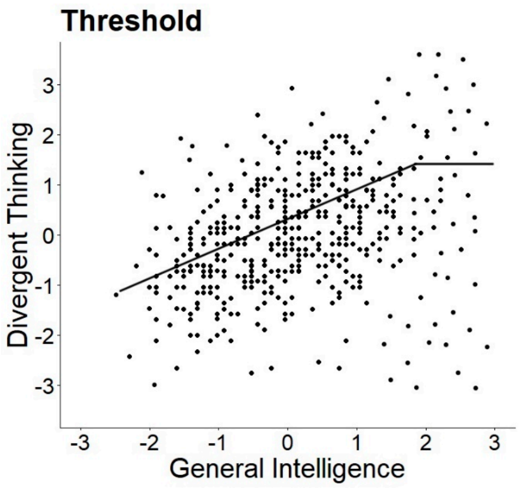


Figure 1. Schematic representations of the relation between creativity and intelligence. The x- and y-axis display z standardized values.

The original formulation of a necessary but not sufficient condition was later (erroneously from many researchers) converted into the so-called threshold hypothesis. The threshold hypothesis states that the relationship between creativity and intelligence varies depending on the level of intelligence. Proponents assume that, below a certain threshold of intelligence, intelligence and creativity show a positive linear relationship, whereas above that threshold intelligence and creativity are uncorrelated (see right plot in Figure 1) or are less strongly correlated. Interestingly, although Guilford is widely named as the originator of the threshold hypothesis, he was no advocate in later publications and theoretically and analytically distinguished between the ideas of assuming a necessary but not sufficient condition, suggesting a threshold. Guilford and colleagues showed in two studies (including 45 tests of divergent production and two IQ tests with various scales) that none of the scatter plots suggested a threshold and that the ubiquitous positive relationship "shows a continuous, gradual shift from low to high IQ", ultimately leading to the completely opposite conclusion that there is no support for any threshold (Guilford and Christensen 1973, p. 251). Guilford and Christensen concluded the absence of a threshold despite a triangular-shaped scatter for most of their plots (e.g., 20 triangular plots for semantic tasks out of 25 tasks), as the linear regression did not show any breaks. This implies that they distinguished intelligence as a necessary but not sufficient condition for being creative (triangular shape of a scatterplot) and the assumption of a threshold given by a difference in correlations between creativity and intelligence tasks at a certain point (break in the regression line; see Figure 1). In summary, there are (at least) three different perspectives on the link between creativity and intelligence: intelligence being (a) a necessary condition, (b) a necessary but not sufficient condition, and (c) the threshold hypothesis. Herein, we overview what researchers understand by the term "partly vary". In the following, we discuss the theoretical assumptions and empirical evidence of the threshold hypothesis.

3. Theoretical Underpinnings of the Threshold Hypothesis

What are the theoretical underpinnings for the threshold hypothesis? Unfortunately, a large amount of research regarding the intelligence-creativity link lacks a thorough theoretical explanation as to why a threshold should exist and if present where it should be (Karwowski and Gralewski 2013). The confusion of terms and the different operationalizations to test the theory might be a direct result of sparse theoretical ideas. However, to discuss where a threshold should exactly be placed in the ability distribution is irrelevant if the "why" is not clear. Although the threshold hypothesis could not be equated with a non-linear relationship between intelligence and creativity, some researchers borrow the theoretical argumentation from other parts of intelligence research, i.e., the ability of the dedifferentiation hypothesis (i.e., Spearman's law of diminishing returns (SLODR, Spearman 1927) or age-related differentiation (Garrett 1946)) to explain the threshold hypothesis of creativity.

At first glance, lending ideas from SLODR seem to be a viable approach, as (general) intelligence directly affects the ability to be creative (e.g., Forthmann et al. 2019; Gilhooly et al. 2007; Silvia et al. 2013). According to SLODR (Spearman 1927), correlations between cognitive abilities decrease with increasing levels of abilities (e.g., Hartung et al. 2018). Transferring this logic would imply that intelligence might facilitate the use of elemental skills (e.g., long-term memory) and, once an advanced level of intelligence is reached, higher levels of intelligence are no longer beneficial for further increasing creative performance, thus leading to a correlational pattern as discussed above. A further example can be found in the differentiation of language ability (Garrett 1946). Initially, it depends on single skills such as oral language comprehension, but the more mature someone gets the more language abilities are differentiated (e.g., reading comprehension, linguistic usage). However, the evidence regarding age differentiation is mixed (Breit et al. 2020; Van Der Maas et al. 2006), and theoretical explanations for this phenomenon are surprisingly sparsely elaborated upon. Some findings support ability differentiation (e.g., Legree et al. 1996), while others use more sophisticated data-analytic approaches to see support for the differentiation hypothesis (Hartung et al. 2018).

However, the consideration of this literature only adds little insight when it comes to why there should be a qualitative gap or threshold in the relation of creativity and intelligence. Moreover, the literature does not provide any cohesive theoretical background for where to set a cutoff a priori. Despite this weak theoretical foundation of the threshold hypothesis, the question if there is a threshold still inspired a considerable amount of studies. In the next paragraph, we summarize the empirical evidence from these studies and give a systematic overview of the findings.

4. Empirical Evaluation of the Threshold Hypothesis

In Table 1, we summarize prominent findings on the threshold hypothesis and give an overview the methods and results of the studies. Strikingly, almost as many different thresholds existed as did studies. The diverse set of results can be attributed to (a) different understandings how the threshold hypothesis is best operationalized, (b) varying sample sizes and sample characteristics, (c) different measures used to assess both intelligence and creative ability, and (d) the analytical procedures to settle a specific threshold.

Although the sample size reported in the studies varied considerably (e.g., N = 88 to N = 12,255), sample size did not seem to affect the results systematically, leaving no evidence for potential publication biases due to missing statistical power. The same was true for other sample features, although some may argue that sample characteristics such as age or ability distribution might influence the results. Age itself had been assumed to affect the factor structure of intelligence as stated in the age differentiation hypothesis (Garrett 1946), but findings were mixed (Breit et al. 2020; Hülür et al. 2011; Tucker-Drob 2009). Moreover, as the threshold was assumed to be at an intelligence score around $z = 1.33$, some samples might have simply failed to include enough cases above that threshold, failing to depict the whole ability spectrum. However, on the contrary, the studies reported in Table 1 show the opposite effect. Studies that oversampled highly gifted participants ($z > 2$, Holling and Kuhn 2008; $z > 1.33$, Preckel et al. 2006) did not find evidence for the threshold hypothesis.

Table 1. Previous investigations in the relation of creativity and intelligence.

Study	Sample	Analytical Method	Measures of Creative Ability (DT)	Measures of Intelligence	Results	Threshold (*z*-Standardized)
Guilford and Christensen (1973)	360 (students)	Scatterplots	10 verbal and figural DT tests [1]	e.g., Stanford Achievement Test	No Threshold	-
Fuchs-Beauchamp et al. (1993)	496 (pre-schoolers)	Correlations in two IQ groups	Thinking Creatively in Action and Movement [2]	e.g., Stanford-Binet Intelligence Scale [8]	Threshold	1.33
Sligh et al. (2005)	88 (college students)	Correlations in two IQ groups	Finke Creative Invention Task [3]	KAIT [9]	No Threshold	-
Preckel et al. (2006)	1328 (students)	Correlations and Multigroup CFA	BIS-HB [4]	BIS-HB [4]	No Threshold	-
Holling and Kuhn (2008)	1070 (students)	Multigroup CFA	BIS-HB [4]	Culture Fair Test [10]	No Threshold	-
Cho et al. (2010)	352 (young adults)	Correlations in two IQ groups	Torrance Test [5]	e.g., WAIS [11]	Threshold	1.33
Jauk et al. (2013)	297 (adults)	SRA	Alternate Uses and Instances [6]	Intelligence-Structure-Battery [12]	Threshold	−1.00 to 1.33
(Karwowski and Gralewski (2013)	921 (students)	Regression analysis and CFA	Test for Creative Thinking-Drawing Production [7]	Raven's Progressive Matrices [13]	Threshold	1.00 to 1.33
Shi et al. (2017)	568 (students)	among others SRA	Torrance Test [5]	Raven's Progressive Matrices [13]	Threshold	0.61 to 1.12

SRA = Segmented Regression Analysis; CFA = Confirmatory Factor Analysis, [1] Wilson et al. (1954); [2] Torrance (1981); [3] Finke (1990); [4] BIS-HB = Berlin Intelligence Structure Test, Jäger et al. (1997); [5] Torrance (1999); [6] Jauk et al. (2013); [7] Urban (2005); [8] Terman and Merill (1973); [9] KAIT = Kaufman Adolescent and Adult Intelligence Test, Kaufman and Kaufman (1993); [10] Cattell and Cattell (1960); [11] WAIS = Wechsler Adult Intelligence Scale, Wechsler (1981); [12] Arendasy et al. (2004); [13] Raven et al. (2003).

Second, the measures used to study the threshold hypothesis might have influenced the results. For example, Jauk et al. (2013) derived varying thresholds for different dimensions of divergent thinking (originality: $z = 0$, creative fluency: $z = 1.33$; ideational fluency: $z = -1.00$), but no threshold for the relation between creative achievement (assessed via self-reports) and intelligence. Overall, the measures of creativity used in the different studies differed largely in breadth and depth of their operationalization (Weiss et al. 2020b). With respect to the measures of intelligence, most studies focused on indicators that assessed fluid intelligence—the ability of abstract reasoning in novel situations—which is an important constituent of overall general intelligence (e.g., Heitz et al. 2005). It is recommended to use a broad measure of creativity when assessing the threshold to eliminate potential item selection bias from narrow tests, although no systematic influence was established (Table 1).

In contrast to the aforementioned study characteristics, the analytical strategy affects whether and where a threshold is found (e.g., Karwowski and Gralewski 2013). Both correlational analyses and segmented regression analyses mostly reported the existence of a threshold (e.g., Cho et al. 2010; Jauk et al. 2013), which varied. Two studies that used correlational analyses confirmed a threshold at $z = 1.33$, despite segmented regression analysis often resulting in different thresholds. Conversely, multi-group confirmatory factor analysis, which evaluates the factor structure (of creativity) in different ability groups, seemed to show no difference between the groups (Holling and Kuhn 2008; Preckel et al. 2006). Based on the previous results, it seemed plausible that the analytical method had a direct impact on the results. Therefore, we considered different methods to probe the threshold hypothesis.

5. Analytical Strategies in the Investigation of the Threshold Hypothesis

Previous studies reported results regarding the threshold hypothesis, most of which were based on a (a) correlational analysis in a split sample, (b) segmented regression analysis, and (c) multi-group confirmatory factor analysis. Additionally, the necessary but not sufficient condition analysis (Dul 2016) has recently gained attention as a statistical tool in the threshold literature. However, the results of the necessary but not sufficient condition analysis could not be directly compared to the results of other methods. Finding a significant proportion above the ceiling did not necessarily imply a threshold (Guilford and Christensen 1973; Ilagan and Patungan 2018), because it did not test for a break in the regression line (see Figure 1). Moreover, there were several open theoretical issues (e.g., causality assumptions that are not examined and further problematized) and issues regarding that method (e.g., no account for sampling error and a high sensitivity to outliers; for a criticism see Ilagan and Patungan 2018). In the present paper, we focused on methods that were used to study the threshold hypothesis rather than the necessary but not sufficient condition.

5.1. Correlational Analysis in Split Sample

The correlational analysis—which often capitalizes on an extreme group design (Preacher et al. 2005)—is the analytical method with the longest tradition in the investigation of the threshold hypothesis (e.g., Cho et al. 2010; Fuchs-Beauchamp et al. 1993; Getzels and Jackson 1962). For this analytical approach, the sample is split into two groups at an a priori set threshold into a low ability group and a high ability group with correlations between intelligence and creativity separately computed. According to the threshold hypothesis, a threshold exists if the correlation is lower or even zero in the high ability group compare to the low ability group (Karwowski and Gralewski 2013). Although this method might seem like a direct translation of the threshold hypothesis into statistical means, it comes with a long list of potential disadvantages. First, the sample split needs a strong theoretical justification for setting the threshold. Given the unclear theoretical roots of the threshold hypothesis, the often-used threshold of $z = 1.33$ is not sufficiently backed up by theory. Eventually, this uncertainty concerning the cutoff yields the risk of exploiting researcher's degrees of freedom (Simmons et al. 2011; Wicherts et al. 2016), probing different thresholds until the desired result is achieved. Second, splitting the sample into two subsamples dichotomizes an otherwise continuous variable (i.e., intelligence), which results in all sorts of statistical problems, such as informational loss, an underestimation of the strength of the bivariate relation, and a mis-categorization of participants that are close to the threshold (MacCallum et al. 2002). Third, as the correlational analysis is based on manifest variables, measurement error and task specificity are not taken into account. Fourth, the analysis most likely suffers from a lack of measurement precision at the more extreme points of the ability distribution because fewer items assess the extremes (Byrne 2010). Fifth, such differences in the correlational patterns in two groups are often biased by samples restricted in dispersion and reliability being lower in the group that is more severely range-restricted. Since the high IQ group in a heterogeneous sample for obvious reasons often contains only a few cases, the parameter estimates (e.g., slope of the regression) are less robust. The point estimate is lower by virtue of variance restriction and by virtue of the fact that item difficulty distribution often follows ability distribution. Therefore, fewer items with good discrimination are available in the tails of the distribution. This indicates that the reliability of person parameters follows the test information function, which is low where few items discriminate. Therefore, sufficient statistical power can often not be reached in extreme groups of small sizes. Consequently, correlational analysis is especially prone to false positive conclusions due to the very nature of the threshold hypothesis.

5.2. Segmented Regression Analysis

Segmented linear regression analysis determines whether different (linear) relationships exist across the continuum of intelligence. This regression analysis includes the estimation of multiple linear models that are fitted for different segments of the data (Ryan and Porth 2007). This means the intelligence continuum is divided several times into two segments and ordinary least squares (OLS) regressions are fitted separately within these segments. A break, which is referred to as a threshold in the linear regression (such as displayed in Figure 1, right panel), means that the slopes of the two regressions differ significantly. A possible advantage of this method is that it can be used to detect a potentially unknown breakpoint rather than confirming an a priori set breakpoint (Ryan and Porth 2007). The method is usually applied if there is a strong theoretical assumption that justifies a break in the relation often in terms of a dose-response relationship (e.g., a critical level of stress leads to preterm birth, Whitehead et al. 2002). Such a strong theoretical basis cannot be assumed in the relation between intelligence and creativity. Furthermore, the segmented regression comes along with several model assumptions that normally distributed and independent residuals are homoscedastic. i.e., OLS regression (Ryan and Porth 2007). However, studies reporting results based on the segmented regression analysis often fail to report tests of homoscedasticity of the data or QQ-plots that examine the normal distribution of residuals. We chose a segmented regression analysis to allow a direct comparison to previous research and because the basic assumptions were met (i.e., homoscedasticity of residuals). Robust alternatives to segmented regression, such as the robust bent line regression (Zhang and Li 2017) or the Robin Hood algorithm for curvilinear relations (Simonsohn 2018), can be considered if the assumptions are violated. These analytical methods assume an unknown change point in a non-linear regression of manifest variables, but the theoretical basis for such an assumption is vague. Moreover, these methods also suffer from problems such as imprecise false positive rates (Type I errors) and the assumption of a change in sign of the regression in two regions (Simonsohn 2018). Besides, methods such as the quantile regression have been applied to investigate thresholds (Dumas 2018; Karwowski et al. 2020), although they do not provide a direct test of a threshold as the segmented regression analysis does.

5.3. Local Structural Equation Modeling

The last analytical method we want to present is a novel approach, termed local structural equation models (LSEM; Hildebrandt et al. 2016). To understand its merits, we will first address the shortcomings of multi-group confirmatory factor analysis (MGCFA), which has been previously used in the threshold literature. MGCFA is a method within the framework of structural equation modeling to analyze measurement parameters (e.g., factor loadings, item intercepts) across different ability groups beyond a simple comparison of correlations (Vandenberg and Lance 2000). Although the latent variable approach is superior compared to simple regressions of manifest variables in an extreme group design, the multi-group setting requires an arbitrary dichotomization of a continuous variable (e.g., z = 2.00, Holling and Kuhn 2008; z = 1.33, Preckel et al. 2006). Another disadvantage of the method is that it does not allow for the direct examination of the correlation of creativity and intelligence, as well as its change across the intelligence continuum. In general, studying the factor variance of creativity over the intelligence continuum might indicate a notable change or threshold (e.g., Holling and Kuhn 2008), i.e., a systematic increase or decrease in factor variance is one way that (de-)differentiation can manifest (Molenaar et al. 2010). However, multi-group confirmatory factor analyses that rely on discretizing a continuous variable at an arbitrary point can mask such a change in the variance. A recent extension of the structural equation models that ameliorates the drawback of an artificial dichotomization of the continuous variable intelligence is LSEM (Hildebrandt et al. 2016). In a nutshell, LSEM involves the fitting of several "conventional" structural equation models along the distribution of a continuous moderator with weighted observations (Olaru et al. 2019). The weight of each observation is based on the proximity of an observation to a specific value of the moderator, so that observations near this focal point provide more information to model estimation than more distant points. In the present context, a series of measurement models for creativity was estimated with intelligence as a continuous moderator. Based on this method, changes in the model fit the factor structure, mean values, and variances without splitting the sample into arbitrary groups (see for example Hartung et al. 2020).

6. The Present Studies

The threshold hypothesis is often attributed to Guilford, though he intended for a necessary but not sufficient condition between intelligence and creativity. In fact, he opposed the idea of a threshold based on empirical findings (1973). Since then, the threshold hypothesis has developed a life of its own, despite the empirical support being weak. In our reading, the theoretical basis of the cognitive mechanisms of the threshold hypothesis, as well as the data analytical approaches, are often not met with the necessary rigor. Applying Occam's razor, no threshold should be assumed or postulated unless convincingly demonstrated otherwise. In the present manuscript, we re-analyzed two data sets that varied with respect to participants' age and the indicators of creativity and intelligence with different analytical strategies. More specifically, we evaluated the relation between intelligence and creativity in both data sets based on the following analytical strategies: (a) scatterplots and heteroscedasticity analysis, (b) segmented regression analysis, and (c) local structural equation models.

7. Method

7.1. Samples and Design

7.1.1. Study 1

The first data set included measures of intelligence, emotional intelligence, and creativity. It was published in the context of investigating the self-other knowledge asymmetry (Neubauer et al. 2018). After data cleaning (excluding n = 6 multivariate outliers with a Mahalanobis distance > 15; Meade and Craig 2012), the total sample included N = 456 adolescents and young adults (ranging from 13 years to 20 years). About 55% of the participants were female. The students were recruited from 13 different public and private schools in rural and urban areas of Austria. For more information please see Neubauer et al. (2018). The dataset is available online via OSF (https://osf.io/v8e5x/).

7.1.2. Study 2

The second data set was part of a larger multivariate study of creativity and its covariates (Goecke et al. 2020; Steger et al. 2020; Weiss et al. 2020a). The analysis was based on $N = 438$ participants after excluding $n = 12$ multivariate outliers with a Mahalanobis distance > 15. Two participants showed high-end performance regarding all creativity indicators. They were not excluded from the data set as they were not flagged as multivariate outliers. The sample included adults between 18 and 49 years. About 65% of the participants were female. For more information regarding the sample and data preparation, see Weiss et al. (2020a). The dataset is available online via OSF (https://osf.io/6fxv5/).

7.2. Measures and Scoring

7.2.1. Study 1

In the first study by Neubauer et al. (2018), intelligence was measured based on the "Intelligenz-Struktur-Analyse" (ISA; Fay et al. 2001), which includes three subtests for verbal, numerical, and spatial reasoning. Creativity was measured using three items from the "Alternate Uses Task" (Jauk et al. 2013). Participants were instructed to name as many original alternate uses for an umbrella, plastic bottle, and a shoe as possible within two minutes. We presented the results for the fluency scoring of answers, i.e., the human coding of the quantity of solutions (for more information, see Neubauer et al. 2018). The fluency scores matched the instruction, which were highly correlated with originality score and frequently applied in the literature. Additionally, we also present the results based on originality scores in the supplementary material (Figures S2–S4).

7.2.2. Study 2

In the second study, intelligence was measured using the verbal and figural subtest of the "Berlin Test of Fluid and Crystallized Intelligence" (Wilhelm et al. 2014). Divergent thinking was measured based on six verbal and figural tests that were either instructed for fluency or originality. The similar attributes test (including 6 items) and the inventing names test (including 18 items) were both adapted from verbal creativity tests (Schoppe 1975). The other two fluency indicators were a typical retrieval fluency test (including 6 items), and the figural fluency test (including 4 items; Jäger et al. 1997). All fluency indicators were rated by humans for the frequency of solutions. Two additional tests (combining objects, French et al. 1963) and inventing nicknames (Schoppe 1975) were rated for the originality/creativity of solutions. Three human raters scored participants' answers on a five-point rating scale (Amabile 1982; Silvia et al. 2008). For more detailed information, please see Weiss et al. (2020a).

7.3. Statistical Analyses

The heteroscedasticity analysis and segmented regression analysis were based on manifest variables. We used *z*-standardized mean values, including either all creativity indicators (Study 1: three items of the Alternate Use Task; Study 2: six tests of fluency and originality) or all intelligence indicators (Study 1: three subtests for verbal, figural, and numerical fluid intelligence; Study 2: indicators for figural and verbal fluid intelligence). The local structural equation modeling relies on a measurement model for creativity using *z*-standardized values. In Study 1, the measurement model was identified using three indicators of the alternate uses task, whereas the model fitted the data well in Study 2 ($\chi^2(9) = 13.31$, $p = 0.15$, CFI = 0.99, RMSEA = 0.03, SRMR = 0.03). In comparison to Weiss et al. (2020a), we modeled creativity as a single factor of divergent thinking, excluding the nested originality factor in the present analysis because it shows low factor saturation and factor variance, which causes estimation problems in LSEM.

7.3.1. Scatterplots and Heteroscedasticity

First, we investigated whether a threshold existed using a scatterplot analysis. Since visual inspection of scatterplots is highly subjective, we tested for heteroscedasticity. Normally distributed residuals indicate homoscedasticity, i.e., the absence of heteroscedasticity. We assumed that if a somehow non-linear relationship between creativity and intelligence existed, values should show a heteroscedasticity, which could be tested with the Breusch–Pagan test (Breusch and Pagan 1979). The Breusch–Pagan test assumes a constant confounding variable variance in the null hypothesis. A non-significant test for heteroscedasticity rendered the existence of a threshold very unlikely.

7.3.2. Segmented Regression Analysis

We used the segmented regression analysis as a second approach to investigate the threshold hypothesis. In this case, a significant change in the slope of the linear regression within the two segments indicated the existence of a threshold. In both studies, intelligence was analyzed as independent variable and divergent thinking as the dependent variable. In addition to estimating the amount and position of possible breakpoints, we used the Davies test to see if any breakpoints occurred between the second greatest and second smallest value (Davies 2002; Muggeo 2008). As no significant changes were assumed if more than 10 segments were specified, we used the recommended default of the Davies test (i.e., 10 segments). If the Davies test was non-significant, the regression parameters were constant across the complete intelligence range.

7.3.3. Local Structural Equation Modeling

Finally, we used LSEM to investigate the threshold hypothesis. In contrast to MGCFA, which relies on the categorization of intelligence as a moderator (e.g., Holling and Kuhn 2008), LSEM allows for the investigation of a factor structure of creativity (Figure 2) across the intelligence continuum. LSEM is a person-sampling method applied to investigate deviations in the measurement model across observations (Olaru et al. 2019). Compared to MGCFA, which requires the grouping of participants, the observations in LSEM are weighted as a function of their proximity to a focal point of intelligence (Hildebrandt et al. 2009). The weights are normally distributed around the focal point, implying a full weight at a focal point and weights decreasing according to the probability density of the normal distribution with increasing distance from the focal point. For example, if the measurement model of divergent thinking (Figure 2) is estimated at the focal point of $z = 1.33$, all participants with an intelligence score of $z = 1.33$ are assigned the highest weight (i.e., 1), and weights decrease as scores are more distant from $z = 1.33$. For each focal point of intelligence, the measurement model of creativity is sequentially estimated based on the weighted samples (Hildebrandt et al. 2016). In Studies 1 and 2, we applied general intelligence as a moderator based on a moderator grid of $z = 0.5$, ranging from $z = -1.50$ to $z = 1.50$, resulting in seven focal points. The effective sample size ranged between $N_{eff} \approx 106$ and $N_{eff} \approx 215$ for Study 1 and $N_{eff} \approx 92$ and $N_{eff} \approx 223$ for Study 2.

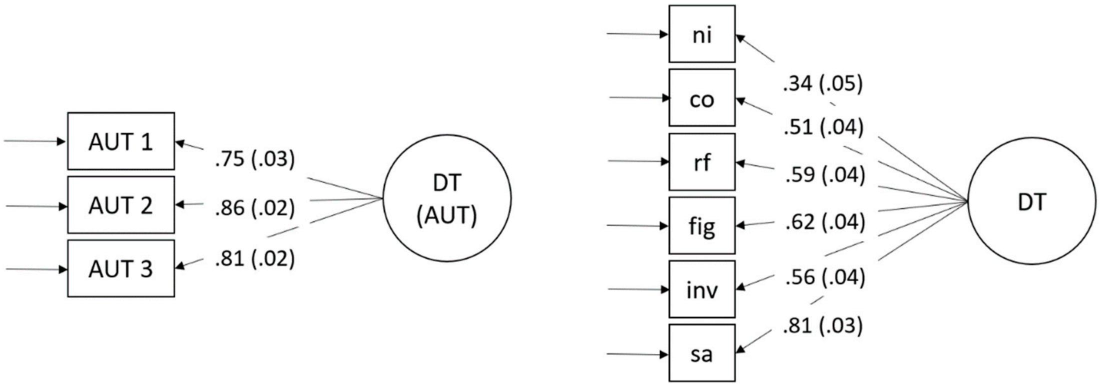


Figure 2. Measurement models for divergent thinking. Study 1 (**left model**), Study 2 (**right model**) including standardized loadings and standard errors. Study 1: AUT are single items of the alternate uses task. Study 2: indicators are test-scores. Fluency test-scores are as follows: sa (similar attributes), in (inventing names), ff (figural fluency), and rf (retrieval fluency). Co (combining objects) and ni (nicknames) are originality indicators that were only instructed and scored for originality.

7.4. Open Science

We conducted all analyses using R version 4.0.2. Segmented regression analyses were estimated using the R package segmented (Muggeo 2008), whereas LSEM was conducted using the packages lavaan and sirt (Robitzsch 2020; Rosseel 2012). To make the present analyses transparent and reproducible, we provided all material (i.e., data set of Study 2, syntax, and supplemental material) at the Open Science Framework. The data set of Study 1 is available online. We also report descriptive statistics (i.e., mean values, standard deviations, and correlations) for the indicators used in the following analysis in the supplementary material (Table S1 and Figure S1).

8. Results

8.1. Scatterplots and Heteroscedasticity

Scatterplots and testing for heteroscedasticity were ur first means to investigate the datasets and to skim for breakpoints in the relation between creativity and intelligence (see Figure 3). In Study 1, the correlation between creativity and intelligence was lower ($r = 0.19$, $p < 0.01$) than in Study 2 ($r = 0.27$, $p < 0.01$). At first glance, the scatterplots (Figure 3, upper part) showed no sign of a threshold. The heteroscedasticity plots (Figure 3, lower part) showed flat lines based on the loess smoothing function, which indicated evenly distributed residuals across the fitted values. Additionally, the Breusch–Pagan test for heteroscedasticity was not significant in both studies (Study 1: BP(1) = 0.64, $p = 0.42$; Study 2: BP(1) = 1.16, $p = 0.28$), so that homoscedasticity could be assumed. The scatterplot and heteroscedasticity plot based on the originality scores (Study 1) are presented in the supplementary material (Figure S2). The Breusch–Pagan test was not significant for originality (BP(1) = 0.47, $p = 0.49$).

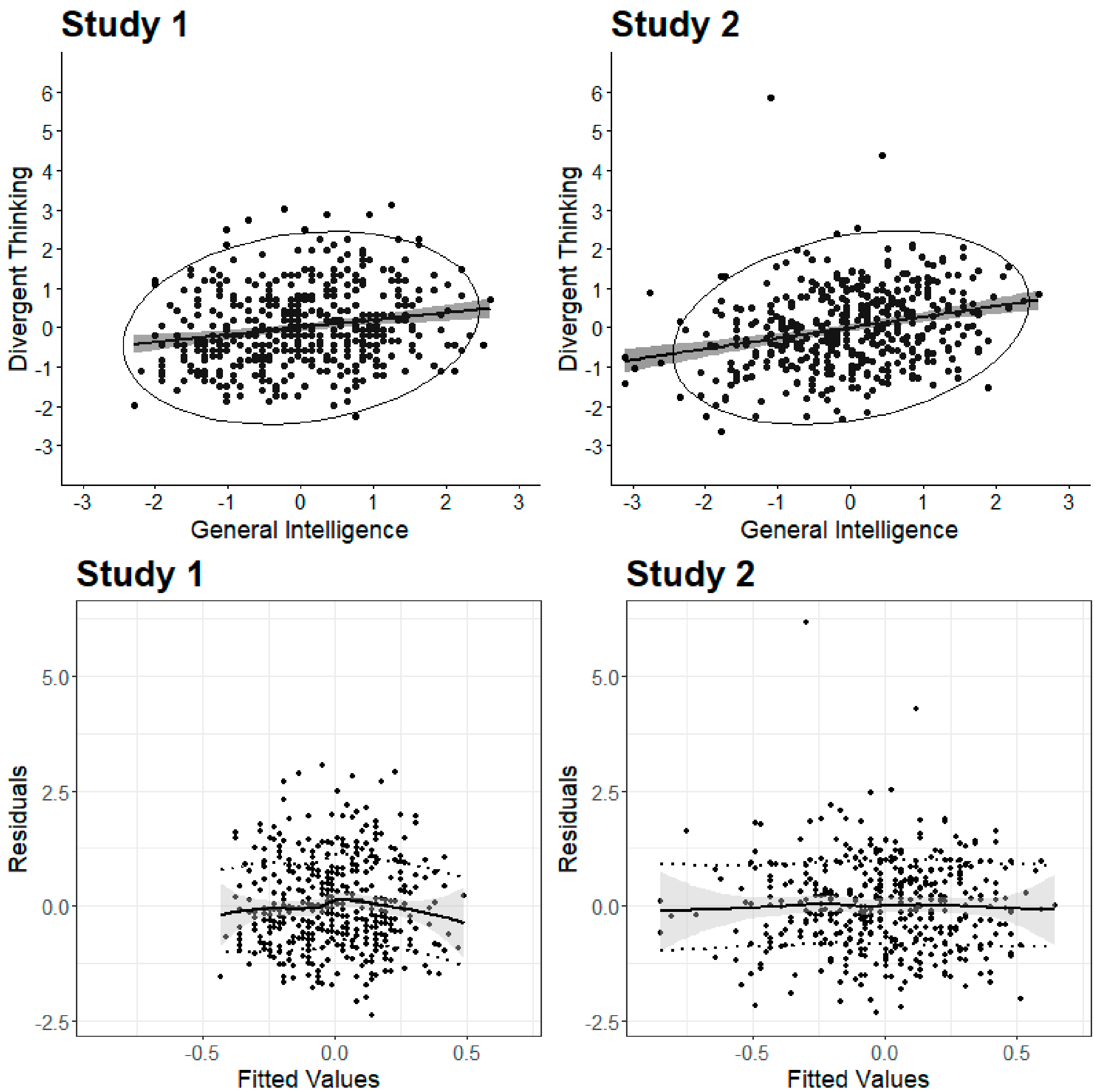


Figure 3. Scatterplots and heteroscedasticity plots. Scatterplots (including the 95% confidence interval) for the correlation between divergent thinking and intelligence are presented upper part. Heteroscedasticity plots including standard errors (grey) and standard deviations of the fitted values (dashed line) are given in the lower part.

8.2. Segmented Regression Analysis

The segmented regression analysis estimates breakpoints in an otherwise linear relationship between two variables. For all breakpoints, the change in slope were not significant; Figure 4 displays the largest change in slopes for Studies 1 and 2. The largest change in slope for the originality indicators (Study 1) is presented in the supplementary material (Figure S3), which was not significant. In sum, there is no evidence for the threshold hypothesis using segmented regression analysis. Nevertheless, we estimated ΔR^2 on Fisher's z-standardized correlation coefficients with z = 1.33 as a breakpoint, because this cutoff was often selected as a potential threshold. In both studies, the number of participants after the breakpoint was small (n_1 = 47, n_2 = 43). The resulting difference was ΔR^2 = 0.06 in Study 1 and ΔR^2 = 0.05 in Study 2.

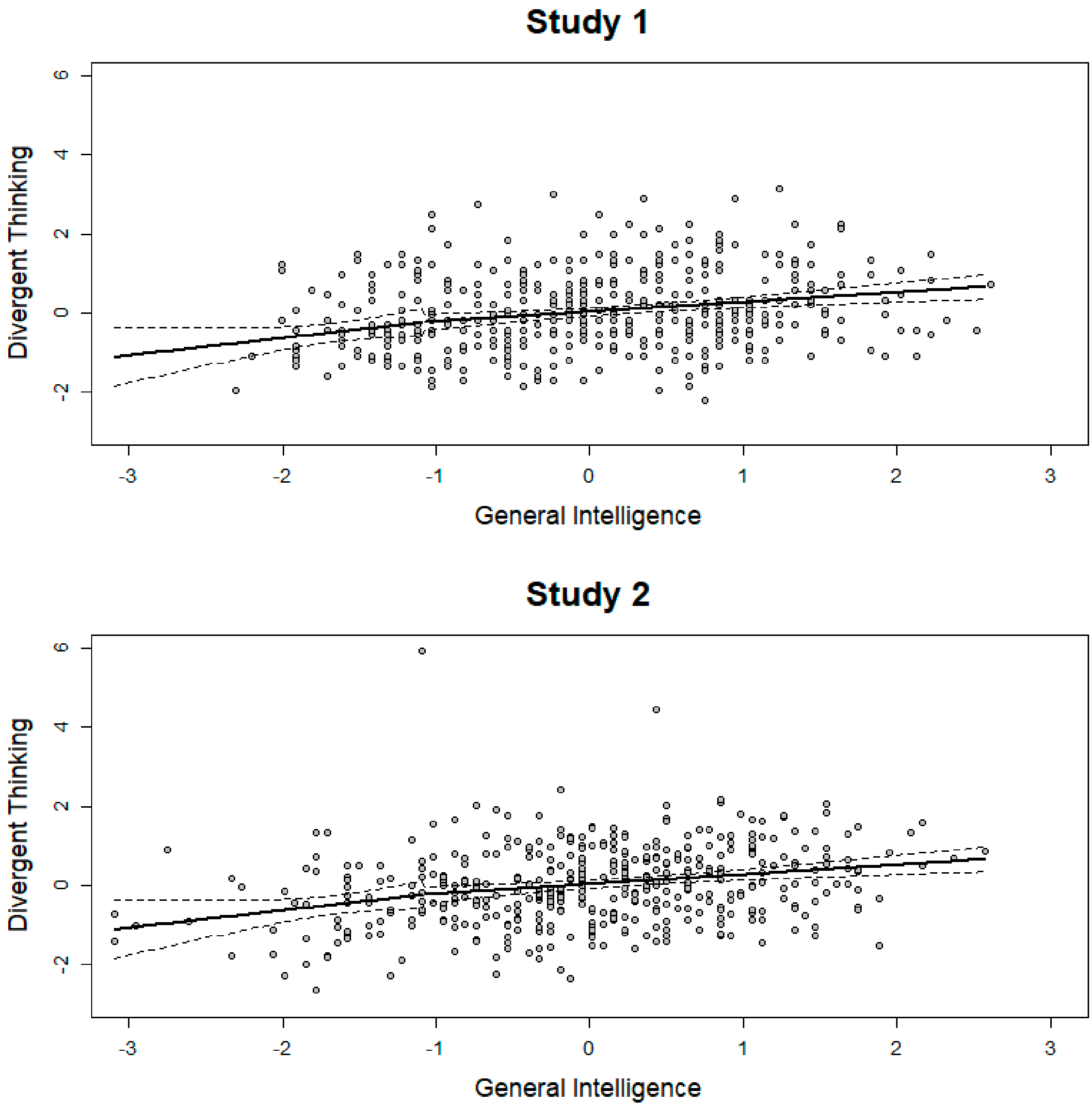


Figure 4. Segmented regression analysis. The breakpoint for the relation between general intelligence and divergent thinking. The dotted line represents the 95% confidence interval.

8.3. Local Structural Equation Models

To detect possible changes in the factor variance of creativity along the intelligence continuum as an indication of a threshold, we fitted local structural equation models in Studies 1 and 2. The model in Study 1 was identified, while the measurement model for creativity fitted well along the intelligence continuum in Study 2, with a slight deterioration in model fit at the tales of the distribution (CFI_{min} = 0.92, $RMSEA_{max}$ = 0.10, and $SRMR_{max}$ = 0.05). No systematic changes in the factor variance of divergent thinking across general intelligence as a moderator were detectable (see Figure 5; see Figure S4 in the supplement for changes in the factor variance of originality). Furthermore, we also fitted a model that constrained the factor loadings to equality to examine if model fit deteriorates. The constraints were introduced to the model with the joint estimation approach for LSEM (separate models at the focal points are equivalently estimated in a multiple group model context; implemented in sirt::lsem.estimate). Similar factor loadings and no decrement in the model fit contradict the idea of a threshold. The loadings at the different focal points in Studies 1 and 2 are displayed in the supplementary material (Study 1: Figure S5; Study 2: Figure S6). As there was no significant change in the model fit, it can be assumed that the loadings do not show greater changes at different focal points in both studies.

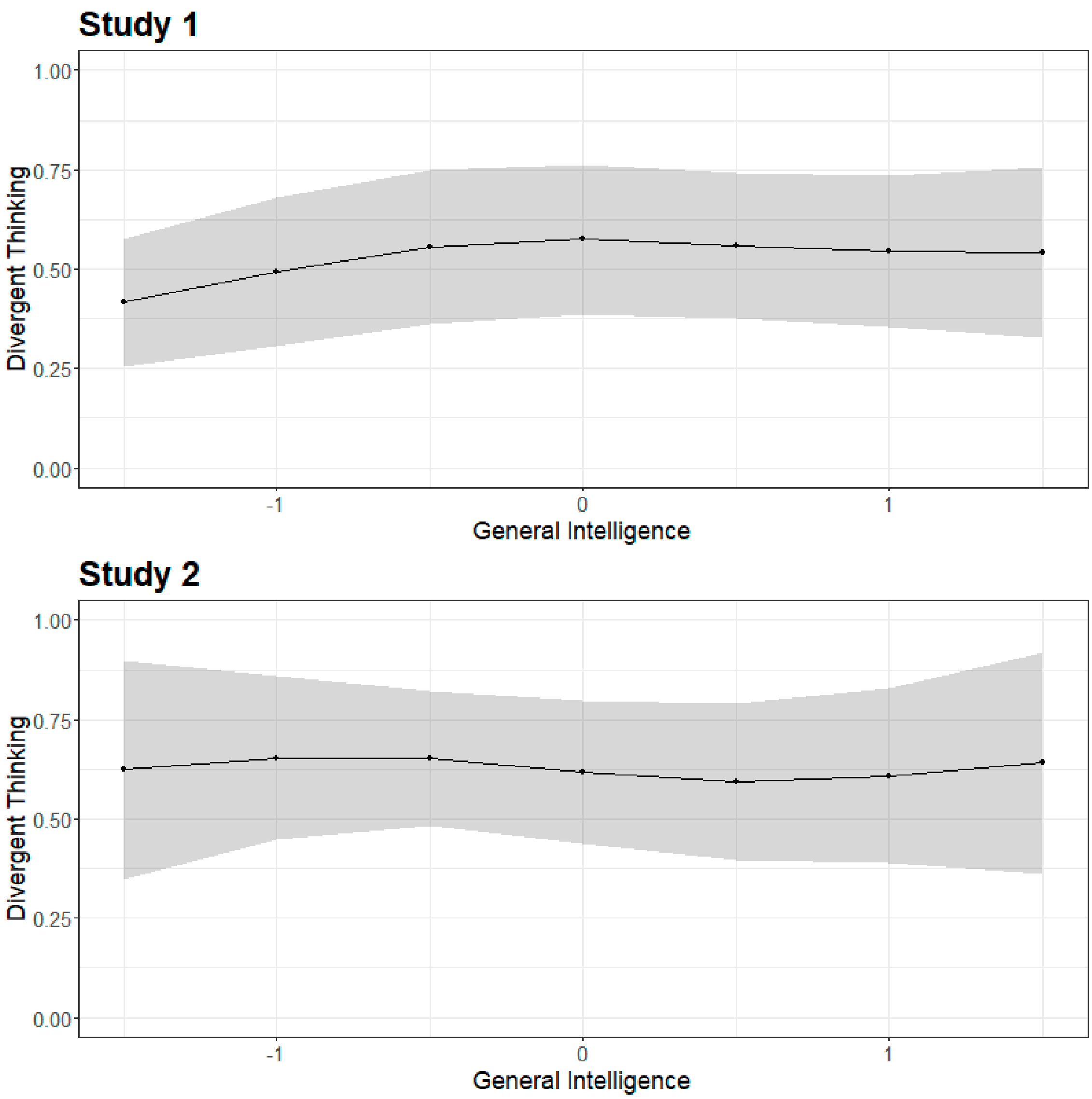


Figure 5. Standardized factor variances at each focal point along the intelligence continuum.

9. Discussion

Investigations regarding a change in the relation between variables above and below a threshold are not limited to creativity research, but can be encountered in many fields such as second language learning (e.g., Cummins 1979). In our reading, these threshold-hypotheses share that they are overgeneralizations of evidence that mainly derived from studies with small sample sizes. Besides, these studies often lacked comprehensive theoretical underpinnings, which is in stark contrast to the extensive attention these hypotheses have attracted over the past decades. Thresholds assumptions should be encountered with some skepticism steered by various conceptual and methodological problems (Takakuwa 2003), and should entail some essential questions.

9.1. Does a Threshold Exist?

In the present case, we reanalyzed two studies with different operationalizations of both fluency/originality and intelligence using three analytical approaches to investigate a potential threshold. Despite these efforts, we were unable to find any compelling evidence for the existence of a threshold. First, the scatterplots of intelligence and creativity did not show any abnormalities and the data were homoscedastic. Second, we found no significant breakpoints using the segmented regression analysis. Finally, the factor variance and factor loadings of a measurement model of creativity did not change across the intelligence continuum. Moreover, since our findings were based on relatively large sample sizes, including different age groups and a variety of different measures of both constructs, we deemed it unlikely that our results were distorted due to a lack of power or sampling issues. This finding is congruent with a number of previous studies that were also unable to find support for an intelligence creativity threshold (e.g., Preckel et al. 2006; Sligh et al. 2005). Remember that Guilford himself led the way when he concluded from two large multivariate studies that he found "no evidence to support a threshold hypothesis regarding the relation of creative potential to IQ" (Guilford and Christensen 1973, p. 252). We concur with this statement. Despite our systematic approach, we did not find any evidence to support the existence of a threshold.

9.2. Why Do Researchers Keep on Finding Evidence Anyway?

While the inference drawn from our results is unambiguous, previous research on the existence of a threshold of creativity and intelligence is not. A narrative review might infer that the results are mixed with some evidence against a threshold (e.g., Holling and Kuhn 2008; Preckel et al. 2006), and some evidence in favor of a threshold (e.g., Jauk et al. 2013; Karwowski et al. 2016). What are potential causes for these inconclusive results? As we saw in the literature review, some differences were caused by the choice of specific analytical approaches, yet the problem goes deeper. Maybe the most apparent is that the threshold is not set a priori. Short of a convincing theory, these cutoffs are arbitrary and leave ample room for many researchers' degrees of freedom in the data analysis. This problem is exacerbated by different handlings of outliers, the choice of analytical tools, etc. (Simmons et al. 2011). Declaring a threshold presupposes its existence and a specific number suggests a precision rarely found in behavioral sciences. As such, it neglects its positivistic identification. Thresholds suggest a qualitative difference of humans below and above the value that is implausible with respect to creativity and intelligence in specific but also, more generally, for psychological dispositions. Even more nuanced approaches—such as the conditional threshold theory (Harris et al. 2019), that supposes that openness plays a critical role in the intelligence-creativity threshold—are adding further complexity and researcher's degrees of freedom. There are additional shortcomings of the prevalent data analytical strategy, such as violated model assumptions (e.g., normally distributed and homoscedastic residuals; Gelman and Hill 2006). In sum, these statistical issues discussed presumably lead to inconsistent results, which have been reported in the literature.

It is important to note that there seems to be a confirmation bias in psychology. This bias usually occurs if subjects are asked to evaluate ambiguous evidence and see their initial expectations confirmed. Equipped with the hypothesis that the relation between intelligence and creativity is weaker above some thresholds, and given the inconclusive literature with partial support for a threshold, researchers are more likely to find that a threshold exists rather than contemplating why their results are at odds with what seems to be a compelling and positive result. Indeed, it is likely that a critical reader suspects those studies that are unable to find a threshold were somewhat flawed, maybe suffering from methodological deficiencies such as small sample sizes, inadequate measures, or other biases. This suspicion is very likely justified since most of the research—including the "positive" findings—suffers from these shortcomings (Ioannidis 2005). In the same vein, researchers who find themselves confronted with a negative result might feel the urge to try searching a little harder to escape these allegations, or to get their results published more easily (Bakker et al. 2012). We are afraid this explanatory bias helps the threshold hypothesis to escape extinction.

9.3. How Should We Approach the Threshold Hypothesis?

We wanted to shed light on a research question that has led to diverging results for over 50 years. We applied analytical strategies that have been used previously—such as the test of heteroscedasticity and the segmented regression analysis—but both approaches usually rely on manifest variables. Therefore, we proposed local structural equation modeling as an additional novel and powerful analytical tool for a continuous treatment of moderators. However, in LSEM, large sample sizes are required to estimate models at each focal point. In the case of the intelligence-creativity threshold hypothesis, this implies that large sample sizes (about $N = 150$; e.g., Muthén and Muthén 2002) are needed at the tails of distribution, which further increases sampling difficulties for normally distributed variables, such as creativity and intelligence. In contrast to other methods, LSEM allows for the detection of non-linear trends and an investigation into the origins of violations of measurement invariance (e.g., Hartung et al. 2020; Olaru et al. 2019).

With the present manuscript, we sought to demonstrate that the search for a specific threshold between intelligence and creativity is a wild goose chase. With that said, we do not want to discourage theoretically well-informed studies that are conducted with the necessary methodological rigor. However, we remain skeptical that a profound theoretical basis exists for further assuming a threshold or a non-linear relationship. In sum, there is no convincing evidence—theoretically or analytically—for the existence of a threshold in the relation between creativity and intelligence. Intelligence is definitely relevant for producing divergent ideas, but its relation appears linear across the continuum of intelligence. If measured broadly, the magnitude of the correlation also seems to fall within an expectable range, which mitigates prior concerns on the strength of the relation between intelligence and creativity. We assume that differentiation will not appear for other factors of creativity (e.g., originality) and intelligence (e.g., crystallized intelligence; Sligh et al. 2005). Nevertheless, studying such aspects in the future—for example, the relation between general retrieval ability, creative retrieval, and crystallized intelligence (e.g., Forthmann et al. 2019), or the overlap between fluency and originality—is interesting to further our understanding of cognitive abilities and its relationship with creativity.

Supplementary Materials: The following are available online at http://www.mdpi.com/2079-3200/8/4/38/s1, Table S1: Descriptive Statistics for all Indicators, Figure S1: Correlations and Bivariate Scatterplots between Manifest Scores, Figure S2: Scatterplot and Heteroscedasticity Plot in Study 1: Originality, Figure S3: Segmented Regression Analysis in Study 1: Originality, Figure S4: Factor Variances at each Focal Point along the Intelligence Continuum in Study 1: Originality, Figure S5: Loadings at the Focal Points in Dataset 1, Figure S6: Loadings at the Focal Points in Dataset 2.

Author Contributions: Conceptualization, S.W., D.S., U.S., and O.W.; Methodology, S.W., D.S., U.S., and O.W.; Validation, D.S., U.S., and O.W; Formal Analysis, S.W.; Investigation, S.W. and D.S.; Resources, O.W. and U.S.; Data Curation, S.W.; Writing—Original Draft Preparation, S.W.; Writing—Review & Editing, D.S., U.S., and O.W.; Visualization, S.W.; Supervision, O.W. and U.S; Project Administration, S.W.; Funding Acquisition, O.W. and U.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank Andrea Hildebrandt and Yadwinder Kaur for recruitment and data collection, as well as helping conceptualizing the creativity measurement. Besides, we thank the numerous research assistants that helped during the data collection as well as with the human coding of creativity measures. We are also thankful for the possibility to analyze the data set published by Neubauer et al. (2018) and we thank Aljoscha Neubauer, Anna Pribil, Alexandra Wallner, and Gabriela Ho-fer for allowing a publication of these results.

Conflicts of Interest: The authors declare no conflict of interest.

References

Amabile, Teresa. M. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology* 43: 997–1013. [CrossRef]

Arendasy, M., L. F. Hornke, M. Sommer, J. Häusler, M. Wagner-Menghin, G. Gittler, and M. Wenzl. 2004. *Manual Intelligence-Structure-Battery (INSBAT)*. Mödling: Schuhfried Gmbh.

Bakker, Marjan, Annette van Dijk, and Jelte M. Wicherts. 2012. The rules of the game called psychological science. *Perspectives on Psychological Science* 7: 543–54. [CrossRef] [PubMed]

Benedek, Mathias, Fabiola Franz, Moritz Heene, and Aljoscha C. Neubauer. 2012. Differential effects of cognitive inhibition and intelligence on creativity. *Personality and Individual Differences* 53: 480–85. [CrossRef] [PubMed]

Breit, Moritz, Martin Brunner, and Franzis Preckel. 2020. General intelligence and specific cognitive abilities in adolescence: Tests of age differentiation, ability differentiation, and their interaction in two large samples. *Developmental Psychology* 56: 364–84. [CrossRef]

Breusch, Trevor S., and Adrian R. Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 5: 1287–94. [CrossRef]

Byrne, Barbara M. 2010. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*, 2nd ed. London: Routledge.

Carroll, John B. 1993. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge: Cambridge University Press.

Cattell, Raymond B., and A. K. S. Cattell. 1960. *Culture Fair Intelligence Test: Scale 2*. Champaign: Institute for Personality and Ability Testing.

Cho, Sun Hee, Jan Te Nijenhuis, Annelies E. M. Vianen, Heui-Baik Kim, and Kun Ho Lee. 2010. The relationship between diverse components of intelligence and creativity. *The Journal of Creative Behavior* 44: 125–37. [CrossRef]

Cropley, Arthur. 2006. In praise of convergent thinking. *Creativity Research Journal* 18: 391–404. [CrossRef]

Cummins, James. 1979. Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research* 49: 222–51. [CrossRef]

Davies, Robert B. 2002. Hypothesis testing when a nuisance parameter is present only under the alternative: Linear model case. *Biometrika* 89: 484–89. [CrossRef]

Diedrich, Jennifer, Emanuel Jauk, Paul J. Silvia, Jeffrey M. Gredlein, Aljoscha C. Neubauer, and Mathias Benedek. 2018. Assessment of real-life creativity: The Inventory of Creative Activities and Achievements (ICAA). *Psychology of Aesthetics, Creativity, and the Arts* 12: 304–16. [CrossRef]

Dul, Jan. 2016. Necessary condition analysis (NCA) logic and methodology of "necessary but not sufficient" causality. *Organizational Research Methods* 19: 10–52. [CrossRef]

Dumas, Denis. 2018. Relational reasoning and divergent thinking: An examination of the threshold hypothesis with quantile regression. *Contemporary Educational Psychology* 53: 1–14. [CrossRef]

Fay, E., G. Trost, and G. Gittler. 2001. *Intelligenz-Struktur-Analyse (ISA)*. Frankfurt: Swets Test Services.

Finke, Ronald A. 1990. *Creative Imagery: Discoveries as Inventions Invisualization*. Mahwah: Lawrence Erlbaum Associates, Inc.

Forthmann, Boris, David Jendryczko, Jana Scharfen, Ruben Kleinkorres, Mathias Benedek, and Heinz Holling. 2019. Creative ideation, broad retrieval ability, and processing speed: A confirmatory study of nested cognitive abilities. *Intelligence* 75: 59–72. [CrossRef]

French, John W., Ruth B. Ekstrom, and Leighton A. Price. 1963. *Manual for Kit of Reference Tests for Cognitive Factors (Revised 1963)*. Princeton: Educational Testing Service.

Fuchs-Beauchamp, Karen D., Merle B. Karnes, and Lawrence J. Johnson. 1993. Creativity and intelligence in preschoolers. *Gifted Child Quarterly* 37: 113–17. [CrossRef]

Garrett, Henry E. 1946. A developmental theory of intelligence. *American Psychologist* 1: 372–78. [CrossRef] [PubMed]

Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/hierarchical Models*. Cambridge: Cambridge University Press.

Getzels, Jacob W., and Philip W. Jackson. 1962. *Creativity and Intelligence: Explorations with Gifted Students*. Hoboken: Wiley.

Gilhooly, K. J., Evie Fioratou, S. H. Anthony, and V. Wynn. 2007. Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology* 98: 611–25. [CrossRef]

Goecke, Benjamin, Selina Weiss, Diana Steger, Ulrich Schroeders, and Oliver Wilhelm. 2020. Its more about what you don't know than what you do know: Perspectives on Overclaiming. *Intelligence* 81. [CrossRef]

Guilford, J. P., and Paul R. Christensen. 1973. The one-way relation between creative potential and IQ*. *The Journal of Creative Behavior* 7: 247–52. [CrossRef]

Guilford, Joy Paul. 1967. *The Nature of Human Intelligence*. New York: McGraw-Hill.

Harris, Alexandra M., Rachel L. Williamson, and Nathan T. Carter. 2019. A conditional threshold hypothesis for creative achievement: On the interaction between intelligence and openness. *Psychology of Aesthetics, Creativity, and the Arts* 13: 322–37. [CrossRef]

Hartung, Johanna, Laura E. Engelhardt, Megan L. Thibodeaux, K. Paige Harden, and Elliot M. Tucker-Drob. 2020. Developmental transformations in the structure of executive functions. *Journal of Experimental Child Psychology* 189: 104681. [CrossRef]

Hartung, Johanna, Philipp Doebler, Ulrich Schroeders, and Oliver Wilhelm. 2018. Dedifferentiation and differentiation of intelligence in adults across age and years of education. *Intelligence* 69: 37–49. [CrossRef]

Heitz, Richard P., Nash Unsworth, and Randall W. Engle. 2005. Working memory capacity, attention control, and fluid intelligence. In *Handbook of Understanding and Measuring Intelligence*. Edited by O. Wilhelm and R. W. Engle. Thousand Oaks: Sage Publications, pp. 61–78.

Hildebrandt, Andrea, Oliver Lüdtke, Alexander Robitzsch, Christopher Sommer, and Oliver Wilhelm. 2016. Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research* 51: 257–58. [CrossRef] [PubMed]

Hildebrandt, Andrea, Oliver Wilhelm, and Alexander Robitzsch. 2009. Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology* 16: 87–102.

Holling, Heinz, and Jörg-Tobias Kuhn. 2008. Does intellectual giftedness affect the factor structure of divergent thinking? Evidence from a MG-MACS analysis. *Psychology Science Quaterly* 50: 283–94.

Hülür, Gizem, Oliver Wilhelm, and Alexander Robitzsch. 2011. Intelligence differentiation in early childhood. *Journal of Individual Differences* 32: 170–79. [CrossRef]

Ilagan, Michael John, and Welfredo Patungan. 2018. The relationship between intelligence and creativity: On methodology for necessity and sufficiency. *Archives of Scientific Psychology* 6: 193–204. [CrossRef]

Ioannidis, John P. A. 2005. Why most published research findings are false. *PLoS Medicine* 2: e124. [CrossRef]

Jäger, A. O., Heinz Martin Süß, and A. Beauducel. 1997. *Berliner Intelligenzstruktur-Test: BIS-Test*. Göttingen: Hogrefe.

Jauk, Emanuel, Mathias Benedek, Beate Dunst, and Aljoscha C. Neubauer. 2013. The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence* 41: 212–21. [CrossRef]

Karwowski, Maciej, and Jacek Gralewski. 2013. Threshold hypothesis: Fact or artifact? *Thinking Skills and Creativity* 8: 25–33. [CrossRef]

Karwowski, Maciej, Dorota M. Jankowska, Arkadiusz Brzeski, Marta Czerwonka, Aleksandra Gajda, Izabela Lebuda, and Ronald A. Beghetto. 2020. Delving into creativity and learning. *Creativity Research Journal* 32: 4–16. [CrossRef]

Karwowski, Maciej, Jan Dul, Jacek Gralewski, Emanuel Jauk, Dorota M. Jankowska, Aleksandra Gajda, Michael H. Chruszczewski, and Mathias Benedek. 2016. Is creativity without intelligence possible? A Necessary Condition Analysis. *Intelligence* 57: 105–17. [CrossRef]

Karwowski, Maciej, Marta Czerwonka, and James C. Kaufman. 2018. Does intelligence strengthen creative metacognition? *Psychology of Aesthetics, Creativity, and the Arts* 14: 353–60. [CrossRef]

Kaufman, Alan S., and Nadeen L. Kaufman. 1993. *The Kaufman Adolescent and Adult Intelligence Test Manual*. Circle Pines: American Guidance Service.

Kim, Kyung Hee. 2005. Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education* 16: 57–66. [CrossRef]

Legree, Peter J., Mark E. Pifer, and Frances C. Grafton. 1996. Correlations among cognitive abilities are lower for higher ability groups. *Intelligence* 23: 45–57. [CrossRef]

MacCallum, Robert C., Shaobo Zhang, Kristopher J. Preacher, and Derek D. Rucker. 2002. On the practice of dichotomization of quantitative variables. *Psychological Methods* 7: 19–40. [CrossRef] [PubMed]

McGrew, Kevin 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. [CrossRef]

Meade, Adam. W., and S. Bartholomew Craig. 2012. Identifying careless responses in survey data. *Psychological Methods* 17: 437–55. [CrossRef]

Molenaar, Dylan, Conor V. Dolan, Jelte M. Wicherts, and Han LJ Van Der Maas. 2010. Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence* 38: 611–24. [CrossRef]

Muggeo, Vito M. R. 2008. Segmented: An R package to fit regression models with broken-line relationships. *R News* 8: 20–25.

Mumford, Michael D., and Tristan McIntosh. 2017. Creative thinking processes: The past and the future. *The Journal of Creative Behavior* 51: 317–22. [CrossRef]

Muthén, Linda K., and Bengt O. Muthén. 2002. How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling* 9: 599–620. [CrossRef]

Neubauer, Aljoscha C., Anna Pribil, Alexandra Wallner, and Gabriela Hofer. 2018. The self–other knowledge asymmetry in cognitive intelligence, emotional intelligence, and creativity. *Heliyon* 4: e01061. [CrossRef] [PubMed]

Nusbaum, Emily C., and Paul J. Silvia. 2011. Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence* 39: 36–45. [CrossRef]

Olaru, Gabriel, Ulrich Schroeders, Johanna Hartung, and Oliver Wilhelm. 2019. Ant colony optimization and local weighted structural equation modeling. A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality* 33: 400–19. [CrossRef]

Preacher, Kristopher J., Derek D. Rucker, Robert C. MacCallum, and W. Alan Nicewander. 2005. Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods* 10: 178–92. [CrossRef]

Preckel, Franzis, Heinz Holling, and Michaela Wiese. 2006. Relationship of intelligence and creativity in gifted and non-gifted students: An investigation of threshold theory. *Personality and Individual Differences* 40: 159–70. [CrossRef]

Raven, John, John C. Raven, and John H. Court. 2003. *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General Overview*. San Antonio: Harcourt Assessment.

Robitzsch, Alexander. 2020. sirt: Supplementary Item Response Theory Models. Available online: https://rdrr.io/github/alexanderrobitzsch/sirt/man/sirt-package (accessed on 20 January 2020).

Rosseel, Yves. 2012. Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software* 48: 1–36. [CrossRef]

Runco, Mark A. 2008. Commentary: Divergent thinking is not synonymous with creativity. *Psychology of Aesthetics, Creativity, and the Arts* 2: 93–96. [CrossRef]

Ryan, Sandra E., and Laurie S. Porth. 2007. *A Tutorial on the Piecewise Regression Approach Applied to Bedload Transport Data*; Washington: US Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Schoppe, Karl-Josef. 1975. *Verbaler Kreativitäts-Test-VKT: ein Verfahren zur Erfassung verbal-produktiver Kreativitätsmerkmale*. Göttingen: Verlag für Psychologie CJ Hogrefe.

Shi, Baoguo, Lijing Wang, Jiahui Yang, Mengpin Zhang, and Li Xu. 2017. Relationship between divergent thinking and intelligence: An empirical study of the threshold hypothesis with Chinese children. *Frontiers in Psychology* 8: 254. [CrossRef]

Silvia, Paul J. 2015. Intelligence and creativity are pretty similar after all. *Educational Psychology Review* 27: 599–606. [CrossRef]

Silvia, Paul J., Beate P. Winterstein, John T. Willse, Christopher M. Barona, Joshua T. Cram, Karl I. Hess, Jenna L. Martinez, and Crystal A. Richard. 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. [CrossRef]

Silvia, Paul J., Roger E. Beaty, and Emily C. Nusbaum. 2013. Verbal fluency and creativity: General and specific contributions of broad retrieval ability (Gr) factors to divergent thinking. *Intelligence* 41: 328–40. [CrossRef]

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22: 1359–66. [CrossRef] [PubMed]

Simonsohn, Uri. 2018. Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science* 1: 538–55. [CrossRef]

Sligh, Allison C., Frances A. Conners, and Beverly Roskos-Ewoldsen. 2005. Relation of creativity to fluid and crystallized intelligence. *The Journal of Creative Behavior* 39: 123–36. [CrossRef]

Spearman, Charles. 1927. *The Abilities of Man*. New York: MacMillan.

Steger, Diana, Ulrich Schroeders, and Oliver Wilhelm. 2020. Caught in the act: Predicting cheating in unproctored knowledge assessment. *Assessment*. [CrossRef]

Süß, Heinz Martin, and André Beauducel. 2005. Faceted models of intelligence. In *Handbook of Understanding and Measuring Intelligence*. Edited by Oliver Wilhelm and Randall W. Engle. Thousand Oaks: Sage Publications.

Takakuwa, Mitsunori. 2003. Lessons from a paradoxical hypothesis: A methodological critique of the threshold hypothesis. In *Előadás: 4th International Symposium on Bilingualism*. Tempe: Arizona State University, p. 12.

Terman, Lewis, and Maud Merill. 1973. *Manual for the Third Revision (Form LM) of the Stanford-Binet Intelligence Scale*. Boston: Houghton Mifflin.

Torrance, E. Paul. 1981. *Thinking Creatively in Action and Movement*. Earth City: Scholastic Testing Service.

Torrance, E. Paul. 1999. *Torrance Tests of Creative Thinking: Thinking Creatively with Pictures, Form A*. Earth City: Scholastic Testing Service.

Tucker-Drob, Elliot M. 2009. Differentiation of cognitive abilities across the life span. *Developmental Psychology* 45: 1097–118. [CrossRef]

Urban, Klaus K. 2005. Assessing creativity: The test for creative thinking—drawing production (TCT-DP). *International Education Journal* 6: 272–80.

Van Der Maas, Han L., Conor V. Dolan, Raoul P. Grasman, Jelte M. Wicherts, Hild M. Huizenga, and Maartje E. Raijmakers. 2006. A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review* 113: 842–61. [CrossRef]

Vandenberg, Robert J., and Charles E. Lance. 2000. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods* 3: 4–70. [CrossRef]

Wechsler, David. 1981. *WAIS Manual; Wechsler Adult Intelligence Scale-Revised*. New York: The Psychological Corporation.

Weiss, Selina, Diana Steger, Kaur Yadwinder, Andrea Hildebrandt, Ulrich Schroeders, and Oliver Wilhelm. 2020a. On the trail of creativity: Dimensionality of divergent thinking and its relation with cognitive abilities and personality. *European Journal of Personality*. [CrossRef]

Weiss, Selina, Oliver Wilhelm, and Patrick Kyllonen. 2020b. A review and taxonomy of creativity measures. *Psychology of Aesthetics, Creativity and the Arts*. submitted.

Whitehead, Nedra, Holly A. Hill, Donna J. Brogan, and Cheryl Blackmore-Prince. 2002. Exploration of threshold analysis in the relation between stressful life events and preterm delivery. *American Journal of Epidemiology* 155: 117–24. [CrossRef] [PubMed]

Wicherts, Jelte M., Coosje L. S. Veldkamp, Hilde E. M. Augusteijn, Marjan Bakker, Robbie C. M. van Aert, and Marcel A. L. M. van Assen. 2016. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology* 7: 1832. [CrossRef]

Wilhelm, Oliver, Ulrich Schroeders, and Stefan Schipolowski. 2014. *Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 8. Bis 10. Jahrgangsstufe [Berlin Test of Fluid and Crystallized Intelligence for Grades 8-10]*. Göttingen: Hogrefe.

Wilson, Robert C., J. P. Guilford, Paul R. Christensen, and Donald J. Lewis. 1954. A factor-analytic study of creative-thinking abilities. *Psychometrika* 19: 297–311. [CrossRef]

Zhang, Feipeng, and Qunhua Li. 2017. Robust bent line regression. *Journal of Statistical Planning and Inference* 185: 41–55. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

VI. Epilogue

Summary

The British Comedian John Cleese stated in his book about creativity that "the first time [he] discovered [he] was a bit creative, it came as a surprise." (Cleese, 2020; p. 11). Discovering your creativity might be a journey full of surprises, however disentangling creativity from a scientific point of view seems to be a journey less surprising, but not less exciting. In the following sections, I summarize the sometimes expected and sometimes surprising results of the theoretical and empirical considerations that are presented above. The first manuscript focuses on the understanding of creativity in psychological but also linguistical research. The second manuscript includes various previously applied measures—in scientific and practical applications—and a taxonomy that allows for a classification of such measures. The third and fourth manuscripts provide empirical evidence for the relation of creativity to other neighboring constructs like cognitive abilities and personality.

Manuscript I: Creativity in Psychological Research versus in Linguistics – Same but different?

The first manuscript (Weiss & Wilhelm, 2020) is mainly inspired by discussing similarities and differences in psychological research and applied linguistics regarding their understanding and application of creativity. With the manuscript, I provide an overview of the linguistic study of creativity. I put the linguistic study into perspective by connecting contentual similarities between linguistics and psychology. Both fields have been studying creativity, although from different angles and based on different terminology. Applied linguistics uses the terminology of *F-* and *E-creativity* (Sampson, 2006) that describe how language can be creatively used either in a fixed space (*F-creativity*) or by enlarging the existing system (*E-creativity*). These terms are linked to fluency, flexibility, and originality (Carroll, 1993). Furthermore, as conceptualized in linguistics, I discuss that creativity might be bound by the nature of rigid language systems that essentially prevent a completely novel contribution (Uhrig, 2020).

Despite this debated austerity, it has been argued that creativity research might benefit from interdisciplinary work between these two disciplines. Although differences in the application of creativity research, both areas have a somewhat similar understanding of what creative is (e.g., novel outcome) and how creativity might emerge (e.g., extending an existing system). Divergent thinking tasks, which are popular in psychological research, assess the originality/novelty of (mostly) verbal utterances. To this end, applied linguistics provides effective standards for considering such originality and novelty in word utterances. Such criteria can be found in the analysis of construction grammar (Hoffmann, 2019) and the latent semantic analysis of answers (Landauer et al., 1998). This conclusion and the plea for interdisciplinary work is the main contribution of this paper. Such interdisciplinary work would further our understanding of creativity and provides new angles for measuring and scoring this construct.

Manuscript II: A Review and Taxonomy of Creativity Measures

In the second manuscript (Weiss et al., under review, submitted 2020), several issues regarding creativity assessment are addressed. Based on this literature review, I aim to provide an overview of creativity measures that have been developed and administered since the early twentieth century, as the literature lacks such a comprehensive collection of measures to date. Besides reviewing the literature, I extrapolate a taxonomy based on integrating various previously suggested taxonomies (e.g., Batey, 2012; Forgeard & Kaufman, 2016; Mumford & Gustafson, 1988; Rhodes, 1961; Said-Metwaly et al., 2017). This taxonomy enables us to categorize measures by several attributes (such as measurement approaches, constructs, data types (Cattell, 1958), prototypical scoring, and psychometric problems). The $N = 213$ measures that are identified are ordered and classified in this taxonomy. The high rater agreement ($\kappa > .83$, mostly $\kappa = 1.00$) between two independent raters confirmes that the taxonomy offers a categorization system, which can be understood easily and applied efficiently. I conclude that the application of divergent thinking measures best approximates the construct of creativity. The review and taxonomy provide a substantial contribution regarding its application for scientific and practical purposes. It comprehensively lists a large selection of creativity tests along with example items and their categorization. This overview can help researchers and practitioners to make an informed decision in choosing measures while at the same time helping to prevent jingle-jangle fallacies (Kelley, 1927). Jingle-jangle fallacies (e.g., different constructs being termed the same labels, or one construct being called different names) are frequent in creativity research, arguably due to the number of measures, vague definitions, and difficult distinction from neighboring constructs. Preventing these fallacies and unifying terminology (see Weiss & Wilhelm, 2020) helps to disentangle the jungle around creativity that has grown in the last century of creativity research.

Manuscript III: On the Trail of Creativity: Dimensionality of Divergent Thinking and its Relation with Cognitive Abilities, Personality, and Insight

The third manuscript (Weiss et al., 2020a) builds upon the results derived from this dissertation's first two manuscripts. In two extensive multivariate studies, divergent thinking measures (N_1 = 152 and N_2 = 298) are applied to understand the nomological net of creativity. Next to the multiple divergent thinking measures, a broad assessment of cognitive abilities (including fluid and crystallized intelligence, mental speed, working memory capacity), personality traits, and insight are included. First, the factor structure of divergent thinking is established by means of confirmatory measurement models. Both studies provide evidence for a general factor of divergent thinking and a nested originality factor. However, the originality factors' variance was limited. The identified measurement model of divergent thinking is then embedded into a larger structural model to investigate the associations of the respective factors with cognitive abilities, personality, and insight. With those studies, I replicate some findings that were suggested by previous studies, such as general cognitive abilities (and specific crystallized intelligence) predict divergent thinking (Benedek et al., 2012), as well as a relation between insight and divergent thinking (Mourgues et al., 2014), and personality (honesty-humility, extraversion) and divergent thinking (Silvia et al., 2011). The finding that openness to experience is not directly correlated with divergent thinking is surprising, given previous studies found both constructs to be moderately related (Puryear et al., 2017). However, in our study, crystallized intelligence mediates the relationship between openness and divergent thinking, which makes sense given the openness factors' nature, including facets such as fantasy, but also intellectual curiosity (Ackerman, 1996; McCrae & Costa, 1989; von Stumm & Ackerman, 2013). The nested originality factor is not meaningfully related to any other constructs, arguably due to its restricted variance. In sum, with this article, several noteworthy results are presented that contribute to understanding creativity and go beyond the findings of previous studies. First, various divergent thinking tasks can be subsumed under one general factor, and a factor that captures the specificity of originality can be established. Second, the nomological net shows that this general factor is predicted by cognitive abilities, some personality factors (extraversion and honesty-humility), and insight. In sum, the covariates explained 32% (Study 1) and 40% (Study 2) of the variance in divergent thinking. Third, originality—despite its great theoretical value—showes limited specificity above and beyond divergent thinking. Fourth, the finding that crystallized intelligence mediated the relationship between openness and divergent thinking challenges previous results that show a moderate relation between openness and divergent thinking. However, most such findings have not controlled for crystallized intelligence (e.g., Silvia et al., 2011).

Manuscript IV: On the Relation between Creativity and Intelligence: An Investigation of the Threshold Hypothesis

In the fourth manuscript (Weiss et al., 2020b), I investigate the relationship between creativity and intelligence. Initially, inspired by Guilford (1967), who stated that intelligence is a *necessary-but-not-sufficient* condition for being creative, a lot of researches claimed that there might be a threshold in the relationship between the two constructs (e.g., Guilford & Christensen, 1973; Holling & Kuhn, 2008; Jauk et al., 2013; Karwowski & Gralewski, 2013). Based on two studies (N_1 = 456 and N_2 = 438), including the total sample presented in manuscript three (Weiss et al., 2020a) and the re-analysis of a second sample (Neubauer et al., 2018), I investigate the existence of such a threshold by applying three different methodological approaches. The manuscript includes methodological approaches that allow a continuous data treatment, such as the analysis of heteroscedasticity, segmented regression analysis, and local structural equation models (e.g., Hildebrandt et al., 2016). All results indicate that empirically no threshold can be assumed. This implies that a non-linear relationship cannot be justified, neither analytically nor theoretically. This paper contributes to a discussion that has been ongoing for over fifty years (Guilford, 1967). I provide a thorough discussion, presenting the lack of a satisfying theoretical basis for such a threshold, and offer guidance regarding methods that should be applied to approaching a possible threshold. In sum, I recommend the application of Occam's razor—postulating parsimony unless proven otherwise—instead of searching for a not further specified kink in a regression.

Interlacing the Empirical Findings

Knowledge regarding conceptual and definitional overlap between different disciplines furthers the understanding of the construct of creativity (Manuscript I). An overview of the enormous variety of various measures helps to prevent jingle-jangle fallacies (Manuscript II). Investigating the relation with neighboring constructs (Manuscript III and IV) broadens our understanding of what creativity might be, but more importantly what it is not. In the following paragraphs, I further embed these findings into the existing literature. These paragraphs include a critical evaluation of previous literature and a representation of the manuscripts' limitations and strengths. Along these lines, I discuss three pressing issues: First, the call for interdisciplinary research, second, the general understanding of creativity and how it is related or drawn apart from what divergent thinking measures, and lastly, the investigation of relationship with other constructs.

First, it seems that interdisciplinary work in creativity research is sparse. However, several articles explicitly stressed the value that interdisciplinary research has for this field. For example, Hennessey and Amabile (2010) have described that seven primary levels for research on creativity (*neurological*, *cognition*, *personality*, *groups*, *social environment*, *culture*, and the *system*) exist. They argue that only interdisciplinary work enables us to understand that whole system (e.g., how social environment affects creative personality). In my opinion, such a holistic view of creativity is overly broad, and it is necessary to understand creativity at a more fine-grained level before we can study it on a holistic level. I present an interdisciplinary work on such a more nuanced level (the understanding of creativity in linguistics and psychology and common individual difference indicators) in the first paper. For example, individual differences lead to different performances in the field of oral poetics. Such different performances can mostly be attributed to individual differences in working memory and retrieval ability (Canovas, 2020; Weiss & Wilhelm, 2020). In sum, there are a large number of disciplines (e.g., sports; Diedrich et al., 2018), and we can only understand how creativity arises in them if we work interdisciplinarily (Sawyer, 1998). Another example is that understanding what marks a creative utterance from the linguistic point (such as construction grammar or the content) can be bundled with the knowledge of why people differ in such performances. Interlacing previous research with results presented in the first manuscript brings out that interdisciplinary work greatly helps us understand what creativity is (e.g., the character of a creative utterance) and how it evolves (e.g., individual differences in vocabulary). However, I argue that interdisciplinary work does not necessarily mean the instant combination of several distinct disciplines, but rather a step by step construction and integration of a broad corpus of knowledge based on the findings of several disciplines.

Second, the general understanding and conceptualization of creativity can and should be problematized. There is a bunch of literature that would argue that whatever we do and whatever efforts we take, we cannot understand and measure "creativity" directly. In line with this, creativity is often presented as an overarching construct that eludes a canonical approach (Runco, 2008), but this is not equitable with all the existing measurement efforts presented across the literature. Such critiques would argue that the divergent thinking tasks that are presented and administered in this dissertation (Manuscript II, III, and IV) are no measures of creativity. In general, there seems to be a trend that some creativity researchers avoid calling the measures they apply "creativity measures". Why is that? We have to ask ourselves, why should a divergent thinking task not be able to grasp this construct's spirit as it is understood in the general research culture? Runco (2008) argued that divergent thinking could not be synonymized with creativity as it just displays a cognitive process that might or might not lead to a creative outcome. I would agree that divergent thinking/production (Guilford, 1967) is an essential prerequisite for producing creative output. However, I would also disagree with Runco (2008) and argue that divergent thinking might be the best and only way to assess creativity. If someone is less skilled in thinking divergently, the possibility of providing creative ideas is arguably low. In clear distinction to that, the question of whether someone behaves creatively is another and might be dependent on motivation, personality, and even chance (Cropley, 2006). This has lead to multiple attempts to assess creativity based on self-evaluations of traits (e.g., personality) that were more or less successful (Diedrich et al., 2018; Forgeard & Kaufman, 2016; Kaufman, 2019). In sum, I argue that divergent thinking is the best approximation of creativity (see III. Manuscript II). I decline the idea that creativity might only be depicted in a higher-order factor, which can never be directly indicated with any tasks whatsoever.

Lastly, I shortly discuss the quest to investigate the relationship of creativity with other relevant constructs. It is necessary to build and enlarge the nomological net of creativity if we want to further our understanding of this construct (Cronbach & Meehl, 1955). That is unless we know what role personality and intelligence play in being creative, we cannot be sure as to what the very nature of creativity is constituted of. However, it appears that some of the previous results presented on this end might be biased by either the measurement (e.g., limited to self-report; Silvia et al., 2011) or debatable analytical procedures due to researchers' degrees of freedom (e.g., Karwowski & Gralewski, 2013). The empirical findings presented in the third and fourth manuscripts might be criticized in terms of the task selection—for the above-described reasons—yet both manuscripts present a broad measurement approach and additionally benefit from sophisticated analytical procedures. In a nutshell, first, the finding that crystallized intelligence mediated the relationship between openness and divergent thinking (Weiss et al., 2020a) might not have been reported before, as other studies have often focused on either personality (Silvia et al., 2011) or intelligence (Silvia et al., 2013), but never put both constructs into consideration at the same time. Second, the results I present regarding the threshold's existence differ from some previously reported results (e.g., Jauk et al., 2013; Karwowski et al., 2016). The literature presented on this kink of relation tends to apply a huge variety of analytical tools (e.g., necessary condition analysis, Karwowski et al., 2016; or various types of regression analysis, Simonsohn, 2019) to find a threshold. I criticize this chase for a regression discontinuity, based on a relatively sparse theoretical background and debatable methodological rigor. However, future work would benefit in general from the discussion of various relational patterns (e.g., theoretically derived change points and different shapes of relationships). Such a discussion should be complemented by the implications that different regressional patterns have for heteroscedasticity and variance.

Where do we go from here?

Despite the profound contributions of the presented manuscripts, the results call for future studies contributing to measurement, scoring, and the nomological net of creativity. In the next paragraph, I present four topics that pursue the research line shown in this dissertation. The first topic outlines and extends ideas of future creativity measurement. The second topic discusses issues of scoring and further research ideas regarding that topic. The third and fourth topic highlights possible extensions of the nomological net of creativity.

Being Creative about Measuring Creativity

As described in the second manuscript, a manifold of various measures that supposedly assess creativity exists. For the reasons discussed in that review paper, I argue that divergent thinking tasks are the measure of choice for evaluating the ability to be creative (Weiss et al., under review, submitted 2020). Despite these tasks' prominence, the measures that assess divergent thinking have not developed much since they were first suggested (Wilson et al., 1954), and the items are often rather uninnovative. Therefore, it can be said that the measures that are applied might not prompt paramount creativity in the participants. For example, when we asked our participants (in the studies presented in manuscript three and four) to name the most original idea that comes to their mind to build a doorstopper at home, about 20% answered that they would fill a sock/towel with rice. Although this answer might be handy, the mere fact that it was the response of 1/5 of the participants makes it doubtful for it to be original. Many other participants used some rather uninspired combinations of household objects (e.g., glue and a box). This means that among over 400 answers, one rarely sees solutions that would be considered (truly) creative, remote, and somehow surprising (e.g., cat and cat food)—maybe even funny or inspirational. The reasons for that might be diverse. One reason could be that the traditionally used items in divergent thinking measures simply do not provoke the participants' above-average creativity. Therefore, I want to offer some ideas that might help in developing future creativity measurements.

The second manuscript already provides a short outlook regarding new measures that rely on technologies that become more and more widespread and accessible, as their economical entry points lower steadily (Thornhill-Miller & Dunpont, 2016). For example, virtual reality devices experience constantly dropping prices, while the technology behind them is getting better year by year. Virtual reality offers various advantages that might help to provoke creative outcomes and lead to divergence in thinking. Virtual realities…

a) …allow a measurement that is close to the real world or explicitly stated in a phantasy setting. This leads to higher ecological validity and, at the same time, to a higher standardization (e.g., by controlling environmental influences, Thornhill-Miller & Dunpont, 2016).
b) …offer the possibility to change characters, concepts, and roles, and therefore comprise different perspectives (Barbot & Kaufman, 2020; Thornhill-Miller & Dunpont, 2016; Zbainos & Lubart, 2016).
c) …lead to the opportunity of interactions with others without being constrained to the people being present in the laboratory (Thornhill-Miller & Dunpont, 2016).

For a first evaluation and validation of such gamified tasks, it might be useful to transfer existing divergent thinking tasks into a virtual setting (e.g., Consequences: *What happens if a pandemic erases all male humans?* The participant finds him-/herself in a virtual post-pandemic space without man and can explore this world (in this case, the virtual space serves as an extended vignette) and indicate further creative actions/futures). This allows for comparing the actions taken or solutions provided in a virtual setting to the conventional test answers. Such environments might be expanded towards new tasks beyond traditional divergent thinking measures (e.g., gamified group interactions). Besides, in the future, the virtual reality setting might be paired with other technologies that help develop a deep and thorough understanding of creativity (e.g., brain stimulation, pharmacological stimulation, or neurofeedback; Thornhill-Miller & Dunpont, 2016).

One recent study applied virtual reality settings to evaluate creativity differences while brainstorming in a real conference room versus brainstorming in a virtual meeting room (e.g., Bourgeois-Bougrine et al., 2020). This study shows that group brainstorming regarding mobility issues in Paris resulted in higher fluency and originality when a virtual setting was used (compared to the standard meeting room). It was concluded that the virtual environment disinhibits individuals with certain traits and abilities (e.g., high risk-taking participants are even more disinhibited and hence more creative; Bourgeois-Bougrine et al., 2020). However, I would argue that the virtual environments (in this example, the virtual meeting room) should be more diverse from the real-life setting or even diverse from what would be possible in a real-life/laboratory setting at all (e.g., meeting an unexplored tribe or escaping from an exit room; maybe being on another and yet unexplored planet). One simply needs to take a glance at the thriving science-fiction genre in the literature, but also on TV, and especially in video games. In sum, technological progress, just like virtual reality, offers great potentials that have yet to be explored and worked out. These technologies are becoming more and more accessible to a broad audience (i.e., home theatres or portable virtual reality devices) and arguably offer a somewhat higher ecological validity than conventional test settings. These technologies will most likely impact how we measure and provoke creative actions in future research. Furthermore, they might also offer new approaches for scoring creativity as much meta-data is produced and gathered in gamified settings. Meta-data can include the actual number of steps a person takes, or might count the number of applications of certain tools a persons can use until a given problem is solved. Times stamps between actions, or creative ability as measured by the ratio of unsuccessful ideas versus successful ideas might offer interesting new accounts about how to approach the measurement of creativity. All these ideas might lead to completely new and innovative ways of measuring and scoring creativity, and I am looking forward to that interesting future.

Considering Alternate Scoring Methods

Until the above described gamified measures in light of virtual realities and the accompanying meta-data concepts are applied, future studies should consider alternative scoring approaches for today's traditional divergent thinking tasks. Therefore, the next pressing avenue for further research tackles the scoring of conventional tasks. As described in the manuscripts presented in this dissertation, the scoring of verbal utterances is required when applying divergent thinking tasks. The human scores that have been used as the gold standard since the invention of these tasks suffer from various problems (e.g., rater effects). Despite the existence of multiple models that can account for differences in raters (e.g., many-facet Rasch models (e.g., Linacre, 1989) or generalized many-facet rater models (e.g., Wang et al., 2014) that add a random effect for the interaction between the item, the rater, and the participant; for an overview see Robitzsch & Steinfeld, 2018), such models often require multiple raters (usually more than three) and are complicated in their application, not to mention uneconomical. These problems have inspired research to apply computerized scorings of creative utterances (see also II. Manuscript I). The basic idea is that the semantic distance, as calculated by latent semantic analysis, between two phrases or between the item and the participants' answer is an indicator of its uncommonness and remoteness (Landauer et al., 1998). The more semantically distant verbal utterances are, the less associative and more remote they are (Wilson et al., 1953). What makes this idea interesting for the scoring of creativity measures is that the semantic distance can be estimated computerized, which is arguably more parsimonious than hiring three or more human raters. In short, the semantic distance is calculated by comparing word vectors (e.g., the creative answers) to matrices of large text corpora that include word distributions and frequencies for an enormous amount of texts (for a detailed description, see: Forthmann et al., 2018). Deriving such computerized estimations of remoteness seems promising and has been successfully applied in research regarding creativity (e.g., Beisemann et al., 2019; Dumas & Dunbar, 2014; Forthmann et al., 2018; Prabhakaran et al., 2014). Nonetheless, the validity of such scores needs further investigation. Future studies must provide evidence if the semantic distance is comparable to the creativity judgments of human raters. Apart from that, several problems should be tackled and eventually eliminated; for example, the data still requires human preparation (e.g., spellcheck) before being applied to the algorithm. Other than that, it must be studied what possible sources might bias the semantic distance. As Forthmann and colleagues (2018) describe, the level of elaboration in an answer can bias this score. Besides, the text corpus' choice might be an additional source of bias, next to the number of missing words (creative creations might not occur in a corpus). In sum, this seems to provide an interesting approach that requires further evaluation and validation.

The latent semantic distance is only useful for the scoring of originality. As stated above, the frequency of provided answers is an essential indicator in evaluating creativity. This leads us to whether we might gather valuable information by combining both fluency and originality scores. There have been previous attempts, such as Simonton's equal odds rule, that study the relation of fluency (quantity) compared to originality (quality) (e.g., Simonton, 2010). The initial idea is that the number of high-quality outcomes (hits, such as real creative inventions) is a linear function of the total efforts taken (fluency) and that can serve as an index of a persons' total productivity (Forthmann et al., 2019b; Simonton, 1977). Therefore, the (classical) equal odds baseline describes the relationship between fluency and originality in terms of a hit ratio. This leads to the (classical) equal odds equation (1) that is expanded by an error term u, as deviations from the strict equal odds (e.g., very creative scientists that are not that fluent) should be considered (Forthmann et al., 2019b): The amount of creative hits of a person j (H_j) is described by their fluency (e.g., the total quantity of ideas) T, and an average hit ratio within the field ρ (under the premise that the ratio of H/T and T are independent; Forthmann et al., 2020):

$$H_j = \rho T_j + u_j \tag{1}$$

The assumption under equal odds has been applied for archival data regarding creative productivity, but it has been rarely applied to divergent thinking measures (Forthmann et al., 2019b, 2020; Jung et al., 2015). The application of the equal odds baseline on divergent thinking tasks intuitively makes sense. Yet, the assumed independence of the hit ratio and the fluency is contrary to the high correlation between originality and fluency often found in divergent thinking tasks (see, for example, Weiss et al., 2020a). This calls for equations and theories that deviate from the original equal odds equation (such as the dual pathway theory that assumes a positive relation between H/T and T; Nijstad et al., 2010). Recent studies have shown that it was dependent on the divergent thinking task administered if the classic equal odds or the assumptions under the dual pathway theory were met (Forthmann et al., 2020). Such task-specificities seem to be counterintuitive and require extensive multivariate studies that compare different models and their assumptions. In sum, new scorings that relate fluency and originality might provide more insight into creative productivity. However, it has to be stated that relying on both scores requires the scoring of a divergent thinking task for both originality and fluency. Scoring originality and fluency in one task is not in line with the instructions, which only allow for one behavior (either fluent or original). Therefore, the models and theories presented above should be expanded towards allowing an application based on different tasks (e.g., relating the fluency in one task with the originality in another task).

The interplay between Creative Fluency, Retrieval Ability, and Working Memory Capacity

Besides establishing new ideas regarding measurement and scoring, it is also important to broaden the nomological net of creativity. As described in the manuscripts, traditionally, creativity has been subordinated to general intelligence in several theoretical attempts (e.g., Jäger et al., 1997; McGrew, 2009). For example, in the *Cattell-Horn-Carroll* model, originality and idea generation are specific abilities subordinated to the second-stratum factor of broad retrieval ability (Carroll, 1993; McGrew, 2009). In addition to that, recent research has stressed the contribution of cognitive abilities (such as general retrieval) to the ability to think divergently (e.g., Forthmann et al., 2019a; Silvia, 2015; Silvia et al., 2013). However, several questions remain unsolved that can be investigated in a more extensive multivariate study: Does creative fluency show significant variance above and beyond tasks of general retrieval ability? Can a factor of originality be extracted beyond fluency and retrieval? And how does working memory capacity affect creative fluency?

Previous studies have shown that a nested factor of creative ideation can be established beyond a general retrieval and beyond a general mental speed factor (Forthmann et al., 2019a) but with restricted factor variance. Therefore, a study comparing several models that only include retrieval ability indicators and divergent thinking (including fluency, flexibility, and originality) helps draw apart how retrieval ability and fluency/originality are intertwined. Other than that, working memory capacity consistently shows an impact on retrieving information (e.g., Rosen & Engle, 1997). For example, people with higher working memory were more flexible in switching between categories (Unsworth et al., 2011). In this light, it is interesting to enlarge the measurement model described above (general retrieval ability and nested divergent thinking factor) by working memory capacity and evaluate its interplay with: a) category switching in retrieval and creative flexibility (Unsworth et al., 2011), b) fluency across more extended periods of time (Beaty & Silvia, 2012), and c) self-reported creative achievements and every-day creativity (Diedrich et al., 2018). A study design, including all these variables and measures, can shed light on various questions that have not been examined together yet.

Does Emotional Creativity exist?

Human cognitive abilities are manifold, and research has not stopped classifying and studying cognitive abilities in the past few decades. This has led to identifying constructs such as emotional intelligence (Mayer & Salovey, 1993). Such a manifold might also be detected in creativity as it is present in our every-day life, and we use it in many different contexts. This leads to the question if the creative handling of emotions and emotional interactions is an ability distinct from creativity as studied so far?

In the following sections, I discuss the distinction between emotional and cognitive creativity as the fourth pressing topic that should be studied to further our understanding of the nomological net of creativity (Cronbach & Meehl, 1955). First, I present the construct definition of emotional creativity and how it can be separated from ordinary creativity. Second, I provide an outlook regarding the measurement of emotional creativity, and third, a study design is presented that allows us to investigate the distinction between the two constructs.

As described above, creativity includes producing novel/original but appropriate ideas (see Weiss et al., under review, submitted 2020; Weiss & Wilhelm, 2020)[1]. Similarly, emotional creativity describes an ability that enables an individual to "experience and express original, appropriate, and authentic combinations of emotions" (Ivcevic et al., 2007, pp. 200). Emotional creativity as a trait describes behaviors that lead to novel and authentic emotions and the understanding of such (Averill, 1999). To my knowledge, there is a lack of studies that rely on multivariate approaches to assess the distinction of cognitive creativity from emotional creativity (e.g., Ivcevic et al., 2007). One study that has investigated this relationship—based

[1] In the following, creativity as described in this definition is referred to as cognitive creativity.

on relatively narrow measurement—showed that the latent correlation between emotional creativity (emotional consequences; Averill & Thomas-Knowles, 1991) and cognitive creativity (cognitive consequences; after Wilson et al., 1954) task is substantial (r = .54), but far from unity (Ivcevic et al., 2007). However, from a Multi-trait-Multi-method point of view (Campbell & Fiske, 1959), it is alarming that the latent correlation between two ability measures of emotional creativity (emotional consequences and emotional triads; Averill & Thomas-Knowles, 1991) is low (r = .24) and the relation with self-reported emotional creativity is even lower (Ivcevic et al., 2007). In sum, the literature investigating emotional creativity is sparse. Previous studies mostly focus on self-report measures, such as the Emotional Creativity Inventory (ECI; Averill, 1999), as shown in a meta-analysis that reported 35 studies using that scale (Kuška et al., 2020). These results call for the development of ability measures of emotional creativity, as Ivcevic and colleagues (2007) defined.

Along these lines, I now present several measures that can serve as emotional creativity measures as an ability. Such measures should distinguish between fluency and originality (e.g., Carroll, 1993), for the reasons described in this dissertation. However, the previous literature is completely lacking such a differentiation when it comes to emotional creativity. I characterize emotional fluency by retrieving emotions that were previously learned or experienced and correctly identifying or using them in a specific situation. I believe that the ability to name emotions fluently is crucial in producing creative emotions. Ability tasks that capture emotional fluency are inspired by typical fluency and retrieval tasks applied in cognitive creativity (e.g., Alternate Uses; Wilson et al., 1954). Emotional fluency tasks can be assessed as follows: (1) listing situations in that a given emotion might occur, (2) naming emotions that another person might feel in a described situation, (3) listing reaction strategies, (4) listing emotions displayed in faces or eyes, and (5) listing emotions that a depicted object can provoke. An example of (1) is "*List as many situations in which you can feel spurred.*". An ability that should be differentiated from fluency is originally dealing with emotions. As described above, I argue that originality should be instructed and scored so that only one single, very original answer is inquired. Originality tasks along this line might include tasks as (1) inventing a unique emotion according to a given situation, (2) producing new emotional composita, (3) describing a creative situation for a given emotion, (4) describing how you can evoke emotions in others in a creative way, (5) combing three emotional words for a given situation (see emotional triad; Averill & Thomas-Knowles, 1991), and (6) creative situational reappraisal. For example, one item for (3) can be "*Describe a situation in which you feel transparent*". Answers regarding that item are "Today someone ran into me because I wasn't visible." or "Everyone knows exactly what I am thinking and knowing. Oh, big brother is watching me." Another example for (4) is "*How would you make people interested?*". Original answers are: "I balance elephants with my feet while the elephants play the trumpet." or "I would let Taylor Swift chug two liters of beer on a slackline in the Alps.". All these tasks are inspired by the results presented in manuscript two and three and they offer an opportunity for a broad and multivariate construct validation of emotional creativity.

Lastly, I present the design of a study that would allow a construct validation of emotional creativity. First, if emotional creativity, measured as ability, can be distinguished from cognitive creativity, needs further examination (Ivcevic et al., 2007). Second, it needs to be studied if emotional creativity measured with the above-described assessments is related to established self-report measures, such as the ECI (Averill, 1999). Therefore, a study design that provides answers to these questions should include a broad assessment of cognitive creativity (including divergent thinking measures presented and discussed in III. Manuscript II; Weiss et al., under review, submitted 2020) and a comprehensive assessment of emotional creativity. As only very few emotional creativity ability tasks exist (see Averill & Thomas-Knowles, 1991), I would include the tasks mentioned earlier, along with existing tasks and self-reported emotional creativity, to ensure a multivariate investigation of emotional creativity. Such a broad assessment in a large sample (> 120) allows for a confirmatory contemplation of emotional and cognitive abilities. In line with the models presented in manuscript three (Weiss et al., 2020a), competing factor models can be used to establish emotional creativity factors and compare such models to a one-factor solution that would assume that only one overarching construct exists. In the next step, the established model can then be related to self-reported emotional creativity. A study, including such a broad construct validation, would significantly contribute to understanding the nomological net of creativity.

Conclusion

De gustibus non est disputandum. Creativity is one of these topics where everyone has an opinion on, and for sure, there are substantial individual differences in taste and the assumption of what it means to be creative. However, scientific contributions help us understand the virtue of creativity, such as measuring creativity and what its predictors are. To this end, I have presented four manuscripts that outlined a) the advantages of an interdisciplinary understanding of creativity (e.g., scoring creative word utterances), b) a taxonomy for classifying existing and future measurements, c) a multivariate study of the nomological net of creativity, and d) the linearity of the relationship between creativity and intelligence. Taken together, the presented manuscripts and the outlook have shown that despite creativity research has now been conducted for over a century, there still remain a lot of unanswered questions.

In this spirit, the most crucial question is: how can we make use of all these scientific results in order to improve our lives? Whom can we help? What future problems can we solve? And how can we foster our children's creativity? The path from research results towards application sometimes appears difficult. However, I believe that creativity research—despite its difficulties to approach a construct that is difficult to grasp—is one of the fields that allows a direct transfer of scientific insights to real-life applications. I appeal, research and applied contexts must work together. For example, the planned analysis and discussion of the creativity data in the upcoming *Programme for International Student Assessment* (PISA) should lead to changes in fostering creativity in schools in the long run, ideally implemented and developed by both scientists and educational stakeholders.

In sum, the understanding we already have gathered, paired with where we go from here, might help us not only to understand what Guilford described in 1950 as the most precious quality of individuals but also how to embed this contemplation into our every-day lives.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*(2), 227–257. https://doi.org/10.1016/S0160-2896(96)90016-1

Averill, J. R. (1999). Individual differences in emotional creativity: Structure and correlates. *Journal of Personality*, *67*(2), 331–371. https://doi.org/10.1111/1467-6494.00058

Averill, J. R., & Thomas-Knowles, C. (1991). Emotional creativity. In K. T. Strongman (Ed.), *International review of studies on emotion* (pp. 269–299). Wiley.

Barbot, B., & Kaufman, J. C. (2020). What makes immersive virtual reality the ultimate empathy machine? Discerning the underlying mechanisms of change. *Computers in Human Behavior*, *111*. https://doi.org/10.1016/j.chb.2020.106431

Batey, M. (2012). The measurement of creativity: From definitional consensus to the introduction of a new heuristic framework. *Creativity Research Journal*, *24*(1), 55–65. https://doi.org/10.1080/10400419.2012.649181

Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(4), 309–319. https://doi.org/10.1037/a0029171

Beisemann, M., Forthmann, B., Bürkner, P.-C., & Holling, H. (2019). Psychometric evaluation of an alternate scoring for the remote associates test. *The Journal of Creative Behavior*, *0*(0), 1-16. https://doi.org/10.1002/jocb.394

Benedek, M., Franz, F., Heene, M., & Neubauer, A. C. (2012). Differential effects of cognitive inhibition and intelligence on creativity. *Personality and Individual Differences*, *53*(4), 480–485. https://doi.org/10.1016/j.paid.2012.04.014

Bourgeois-Bougrine, S., Richard, P., Burkhardt, J.-M., Frantz, B., & Lubart, T. (2020). The expression of users' creative potential in virtual and real environments: An exploratory study. *Creativity Research Journal*, *32*(1), 55–65. https://doi.org/10.1080/10400419.2020.1712162

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Canovas, C. P. (2020). Learning formulaic creativity: Chunking in verbal art and speech. *Cognitive Semiotics 13*(1). https://doi.org/10.1515/cogsem-2020-2023

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. University Press.

Cattell, R. B. (1958). What is" objective" in" objective personality tests?". *Journal of Counseling Psychology*, *5*(4), 285–289. https://doi.org10.1037/h0046268

Cleese, J. (2020). *Creativity—A short and cheerful guide*. Penguin Random House LLC.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal*, *18*(3), 391–404. https://doi.org/10.1207/s15326934crj1803_13

Diedrich, J., Jauk, E., Silvia, P. J., Gredlein, J. M., Neubauer, A. C., & Benedek, M. (2018). Assessment of real-life creativity: The inventory of creative activities and achievements (ICAA). *Psychology of Aesthetics, Creativity, and the Arts*, *12*(3), 304–316. 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/10.1016/j.tsc.2014.09.003

Forgeard, M. J. C., & Kaufman, J. C. (2016). Who cares about imagination, creativity, and innovation, and why? A review. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(3), 250–269. https://doi.org/10.1037/aca0000042

Forthmann, B., Oyebade, O., Ojo, A., Günther, F., & Holling, H. (2018). Application of latent semantic analysis to divergent thinking is biased by elaboration. *The Journal of Creative Behavior 53*(4), 559-575. https://doi.org/10.1002/jocb.240

Forthmann, B., Jendryczko, D., Scharfen, J., Kleinkorres, R., Benedek, M., & Holling, H. (2019a). 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., Szardenings, C., & Dumas, D. (2019b). Testing equal odds in creativity research. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000294

Forthmann, B., Szardenings, C., Dumas, D., & Feist, G. J. (2020). Strict equal odds: A useful reference to study the relationship between quality and quantity. *Creativity Research Journal*, *0*(0), 1–10. https://doi.org/10.1080/10400419.2020.1827605

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

Guilford, J. P., & Christensen, P. R. (1973). The one-way relation between creative potential and IQ*. *The Journal of Creative Behavior*, *7*(4), 247–252. https://doi.org/10.1002/j.2162-6057.1973.tb01096.x

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, *61*(1), 569–598. https://doi.org/10.1146/annurev.psych.093008.100416

Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research*, *51*(2–3), 257–258. https://doi.org/10.1080/00273171.2016.1142856

Hoffmann, T. (2019). Language and creativity: A Construction Grammar approach to linguistic creativity. *Linguistics Vanguard*, *5*(1), 1-8. https://doi.org/10.1515/lingvan-2019-0019

Holling, H., & Kuhn, J.-T. (2008). Does intellectual giftedness affect the factor structure of divergent thinking? Evidence from a MG-MACS analysis. *Psychology Science Quarterly*, *50*(2), 283–294.

Ivcevic, Z., Brackett, M. A., & Mayer, J. D. (2007). Emotional intelligence and emotional creativity. *Journal of Personality*, *75*(2), 199–236. https://doi.org/10.1111/j.1467-6494.2007.00437.x

Jäger, A. O., Süß, H. M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: BIS-Test*. Hogrefe.

Jauk, E., Benedek, M., Dunst, B., & Neubauer, A. C. (2013). The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence*, *41*(4), 212–221. https://doi.org/10.1016/j.intell.2013.03.003

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*, 1-8. https://doi.org/10.3389/fpsyg.2015.00864

Karwowski, M., & Gralewski, J. (2013). Threshold hypothesis: Fact or artifact? *Thinking Skills and Creativity*, *8*, 25–33. https://doi.org/10.1016/j.tsc.2012.05.003

Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D. M., Gajda, A., Chruszczewski, M. H., & Benedek, M. (2016). Is creativity without intelligence possible? A Necessary

Condition Analysis. *Intelligence*, *57*, 105–117. https://doi.org/10.1016/j.intell.2016.04.006

Kaufman, J. C. (2019). Self-assessments of creativity: Not ideal, but better than you think. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 187–192. https://doi.org/10.1037/aca0000217

Kelley, T. L. (1927). *Interpretation of Educational Measurements*. World Book.

Kuška, M., Trnka, R., Mana, J., & Nikolai, T. (2020). Emotional creativity: A meta-analysis and integrative review. *Creativity Research Journal*, *32*(2), 151–160. https://doi.org/10.1080/10400419.2020.1751541

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Linacre, J. (1989). *Many-facet Rasch measurement*. MESA Press.

Mayer, J. D., & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence*, *17*(4), 433–442. https://doi.org/10.1016/0160-2896(93)90010-3

McCrae, R. R., & Costa, P. T. (1989). More reasons to adopt the Five-Factor Model. *American Psychologist*, *44*(2), 451–452. https://doi.org/10.1037/0003-066X.44.2.451

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), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Mourgues, C. V., Preiss, D. D., & Grigorenko, E. L. (2014). Reading skills, creativity, and insight: Exploring the connections. *The Spanish Journal of Psychology*, *17*(58), 1-10. https://doi.org/10.1017/sjp.2014.59

Mumford, M. D., & Gustafson, S. B. (1988). Creative Syndrome: Integration, Application, and Innovation. *Psychological Bulletin*, *103*(1), 27–43. https://doi.org/10.1037/0033-2909.103.1.27

Neubauer, A. C., Pribil, A., Wallner, A., & Hofer, G. (2018). The self–other knowledge asymmetry in cognitive intelligence, emotional intelligence, and creativity. *Heliyon*, *4*(12), e01061. https://doi.org/10.1016/j.heliyon.2018.e01061

Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology*, *21*(1), 34–77. https://doi.org/10.1080/10463281003765323

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*(3), 641–659. https://doi.org/10.3758/s13428-013-0401-7

Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relating personality and creativity: considering what and how we measure. *The Journal of Creative Behavior, 53*(2), 232-245. https://doi.org/10.1002/jocb.174

Rhodes, M. (1961). An Analysis of Creativity. *The Phi Delta Kappan*, *42*(7), 305–310.

Robitzsch, A., & Steinfeld, J. (2018). Item response models for human ratings: Overview, estimation methods, and implementation in R. *Psychological Test and Assessment Modeling*, *60*(1), 101–139.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Jorunal of Experimental Psychology: General*, *126*(3), 211–227. https://doi.org/10.1037/0096-3445.126.3.211

Runco, M. A. (2008). Commentary: Divergent thinking is not synonymous with creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 93–96. https://doi.org/10.1037/1931-3896.2.2.93

Said-Metwaly, S., Noortgate, W. V. den, & Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. *Creativity. Theories – Research - Applications*, *4*(2), 238–275. https://doi.org/10.1515/ctra-2017-0013

Sampson, G. (2006). Two ideas of creativity. In M. Hinton (Eds.), *Evidence, Experiment and Argument in Linguistics and Philosophy of Language*, (pp. 15-26). Peter Lang Edition.

Sawyer, R. K. (1998). The interdisciplinary study of creativity in performance. *Creativity Research Journal*, *11*(1), 11–19. https://doi.org/10.1207/s15326934crj1101_2

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

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*(5), 328–340. https://doi.org/10.1016/j.intell.2013.05.004

Silvia, P. J. (2015). Intelligence and creativity are pretty similar after all. *Educational Psychology Review*, *27*(4), 599–606. https://doi.org/10.1007/s10648-015-9299-1

Simonsohn, U. (2019). Corrigendum: Two lines: A valid alternative to the invalid testing of U-Shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science*, *2*(4), 410–411. https://doi.org/10.1177/2515245919894972

Simonton, D. K. (1977). Creative productivity, age, and stress: A biographical time-series analysis of 10 classical composers. *Journal of Personality and Social Psychology*, *35*(11), 791–804. https://doi.org/10.1037/0022-3514.35.11.791

Simonton, D. K. (2010). Creative thought as blind-variation and selective-retention: Combinatorial models of exceptional creativity. *Physics of Life Reviews*, *7*(2), 156–179. https://doi.org/10.1016/j.plrev.2010.02.002

Thornhill-Miller, B., & Dunpont, J. M. (2016). Virtual reality and the enhancement of creativity and innovation: Under recognized potential among converging technologies? *Journal of Cognitive Education and Psychology*, *15*(1), 102–121. https://doi.org/10.1891/1945-8959.15.1.102

Uhrig, P. (2020). Creative Intentions – The fine line between ‘creative’ and ‘wrong.’ *Cognitive Semiotics 13*(1). https://doi.org/10.1515/cogsem-2020-2027

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2011). Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *Quarterly Journal of Experimental Psychology*, *64*(3), 447–466. https://doi.org/10.1080/17470218.2010.505292

von Stumm, S., & Ackerman, P. L. (2013). Investment and intellect: A review and meta-analysis. *Psychological Bulletin*, *139*(4), 841–869. https://doi.org/10.1037/a0030746

Wang, W.-C., Su, C.-M., & Qiu, X.-L. (2014). Item response models for local dependence among multiple ratings. *Journal of Educational Measurement*, *51*(3), 260–280. https://doi.org/10.1111/jedm.12045

Weiss, S., Steger, D., Kaur, Y., Hildebrandt, A., Schroeders, U., & Wilhelm, O. (2020a). On the trail of creativity: Dimensionality of divergent thinking and its relation with cognitive abilities and personality. *European Journal of Personality.* https://doi.org/10.1002/per.2288

Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020b). A reappraisal of the threshold hypothesis of creativity and intelligence. *Journal of Intelligence*, *8*(4), 38-58. https://doi.org/10.3390/jintelligence8040038

Weiss, S., Wilhelm, O., & Kyllonen, P. (under review, submitted 2020). A review and taxonomy of creativity measures. *Journal of Aesthetics, Creativity, and the Arts.*

Weiss, S., & Wilhelm, O. (2020). Coda: Creativity in psychological research versus in linguistics – Same but different? *Cognitive Semiotics*, *13*(1), 1-15. https://doi.org/10.1515/cogsem-2020-2029

Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). The measurement of individual differences in originality. *Psychological Bulletin*, *50*(5), 362-370. https://doi.org/10.1037/h0060857

Wilson, R. C., Guilford, J. P., Christensen, P. R., & Lewis, D. J. (1954). A factor-analytic study of creative-thinking abilities. *Psychometrika*, *19*(4), 297–311. https://doi.org/10.1007/BF02289230

Zbainos, D., & Lubart, T. I. (2016). Cognitive education and creative cognition. *Journal of Cognitive Education and Psychology*, *15*(1), 3–5. https://doi.org/DOI:10.1891/1945-8959.15.1.3

VII. Appendix

Manuscript III: On the Trail of Creativity: Dimensionality of Divergent Thinking and Its Relation With Cognitive Abilities, Personality and Insight

Supplementary Material

Published Online in the Open Science Framework – Center for Open Science:

Table SM 1

Descriptive Statistics for all Creativity Indicators Including Inter-Rater Reliabilities (ICCs) for Study 1

Task	Item	*N*	*Mean* (*SD*)	*ICC*
SA	1	144	10.47 (4.15)	.98
	2	144	4.31 (2.17)	.98
	3	144	4.43 (2.93)	1.00
	4	144	11.40 (3.12)	.99
	5	144	3.93 (2.38)	.99
	6	144	5.22 (2.40)	.97
IN	1	147	0.95 (0.69)	.97
	2	147	0.53 (0.52)	.95
	3	147	0.96 (0.66)	.98
	4	147	0.51 (0.53)	.95
	5	147	0.98 (0.71)	.98
	6	147	0.64 (0.55)	.97
	7	147	1.02 (0.81)	.99
	8	147	0.49 (0.52)	.96
	9	147	1.07 (0.67)	.97
	10	147	0.51 (0.52)	.97
	11	147	0.89 (0.64)	.97
	12	147	0.68 (0.54)	.97
	13	147	1.06 (0.70)	.97
	14	147	0.44 (0.51)	.98
	15	147	1.12 (0.69)	.97
	16	147	0.54 (0.54)	.97
	17	147	0.98 (0.76)	.99
	18	147	0.58 (0.54)	.97
FF	1	149	4.34 (1.53)	.92
	2	149	5.87 (1.85)	1.00
	3	149	5.47 (2.29)	.97
	4	149	5.28 (2.19)	.99

RF	1	145	11.36 (4.65)	1.00
	2	145	7.18 (2.83)	1.00
	3	145	10.76 (4.42)	1.00
	4	145	5.99 (2.57)	.98
	5	145	4.93 (2.05)	.99
	6	145	9.56 (3.21)	1.00
CO	1	146	2.10 (1.09)	.58
	2	146	1.91 (1.04)	.83
	3	146	1.72 (1.02)	.81
	4	146	1.39 (1.04)	.77
	5	146	2.02 (0.81)	.82
	6	146	1.94 (1.00)	.77
	7	146	1.54 (0.43)	.68
	8	146	1.63 (0.85)	.82
	9	146	1.49 (0.67)	.61
	10	146	1.88 (0.94)	.86
	11	146	1.90 (0.89)	.80
	12	146	2.07 (0.64)	.80
NI	1	147	1.77 (1.11)	.88
	2	147	2.35 (1.16)	.85
	3	147	2.14 (1.16)	.85
	4	147	2.19 (1.03)	.81
	5	147	2.21 (1.15)	.85
	6	147	1.93 (1.25)	.90
	7	147	1.92 (1.07)	.86
	8	147	1.94 (1.14)	.87
	9	147	1.93 (1.28)	.90

Note. Fluency indicators are SA (similar attributes), IN (inventing names), FF (figural fluency), and RF (retrieval fluency). CO (combining objects) and NI (nicknames) are indicators of Originality that were only instructed for originality. Fluency indicators were rated regarding the quantity of correct responses, originality was rated regarding the quality of a single given response.

Table SM 2

Descriptive Statistics for all Creativity Indicators Including Inter-Rater Reliabilities (ICCs) for Study 2

Task	Item	*N*	*Mean* (*SD*)	*ICC*
SA	1	292	10.73 (4.14)	.99
	2	292	5.18 (2.60)	.99
	3	292	5.24 (2.57)	1.00
	4	292	12.26 (3.41)	.99
	5	292	5.45 (2.57)	1.00
	6	292	5.56 (2.31)	.90
IN	1	290	0.87 (0.63)	.98
	2	290	0.63 (0.54)	.97
	3	290	0.95 (0.65)	.97
	4	290	0.63 (0.55)	.97
	5	290	1.07 (0.65)	.98
	6	290	0.64 (0.56)	.97
	7	290	1.07 (0.73)	.99
	8	290	0.59 (0.53)	.96
	9	290	1.03 (0.63)	.98
	10	290	0.63 (0.56)	.96
	11	290	0.94 (0.61)	.97
	12	290	0.74 (0.56)	.97
	13	290	1.03 (0.63)	.98
	14	290	0.46 (0.52)	.97
	15	290	1.15 (0.68)	.99
	16	290	0.67 (0.54)	.98
	17	290	0.97 (0.65)	.97
	18	290	0.59 (0.52)	.97
FF	1	293	4.49 (1.82)	.96
	2	293	6.62 (2.39)	.99
	3	293	6.13 (2.87)	1.00
	4	293	6.10 (2.43)	.98
RF	1	284	10.70 (5.21)	.99
	2	284	7.40 (3.19)	.99
	3	284	10.19 (4.18)	1.00
	4	284	6.53 (2.23)	.97
	5	284	5.61 (1.99)	.99
	6	284	10.06 (3.56)	1.00
CO	1	293	2.10 (0.97)	.53
	2	293	1.86 (1.07)	.87
	3	293	1.72 (1.00)	.78
	4	293	1.35 (0.96)	.76

	5	293	1.95 (0.86)	.84
	6	293	1.82 (1.02)	.79
	7	293	1.50 (0.51)	.82
	8	293	1.55 (0.94)	.87
	9	293	1.39 (0.54)	.65
	10	293	2.05 (0.86)	.83
	11	293	1.84 (0.92)	.85
	12	293	2.18 (0.53)	.79
NI	1	293	1.76 (1.04)	.84
	2	293	2.19 (1.10)	.87
	3	293	1.87 (1.11)	.86
	4	293	1.92 (1.13)	.86
	5	293	2.12 (1.18)	.87
	6	293	1.84 (1.09)	.85
	7	293	1.94 (1.14)	.90
	8	293	1.82 (1.06)	.85
	9	293	1.72 (1.11)	.86

Note. Fluency indicators are SA (similar attributes), IN (inventing names), FF (figural fluency), and RF (retrieval fluency). CO (combining objects) and NI (nicknames) are indicators of Originality that were only instructed for originality. Fluency indicators were rated regarding the quantity of correct responses, originality was rated regarding the quality of a single given response.

Table SM 3

Descriptive Statistics for all Insight Indicators Including Inter-Rater Reliabilities (ICCs) for Study 2

Task	Item	*N*	*Mean* (*SD*)	*ICC*
ANAorg	1	297	0.77 (0.58)	.97
	2	297	1.02 (0.75)	.98
	3	297	0.77 (0.60)	.96
	4	297	0.64 (0.87)	.99
	5	297	0.92 (0.77)	.95
	6	297	1.03 (0.86)	.90
	7	297	0.44 (0.73)	.98
	8	297	0.61 (0.68)	.95
	9	297	0.49 (0.73)	.93
	10	297	0.73 (0.54)	.89
	11	297	0.58 (0.66)	.94
	12	297	0.86 (0.72)	.98
	13	297	0.43 (0.70)	.91
	14	297	0.10 (0.45)	.95
	15	297	0.05 (0.25)	.75
	16	297	0.70 (0.62)	.97
	17	297	0.12 (0.55)	.97
	18	297	0.29 (0.51)	.99
SCRorg	1	297	1.61 (1.15)	.83
	2	297	1.58 (1.19)	.86
	3	297	1.72 (0.97)	.75
	4	297	1.99 (1.14)	.74
	5	297	1.53 (0.94)	.59
	6	297	2.16 (1.24)	.83
	7	297	1.71 (1.13)	.77
	8	297	1.79 (1.07)	.81
	9	297	1.70 (1.13)	.82
	10	297	1.87 (1.12)	.85
	11	297	1.69 (1.21)	.82
	12	297	1.89 (1.01)	.80
	13	297	1.55 (0.79)	.67
	14	297	1.89 (1.46)	.89
SCRflu	1	297	1.75 (1.11)	.97
	2	297	3.38 (1.48)	.93
	3	297	3.86 (1.35)	.91
	4	297	3.11 (1.64)	.95
	5	297	1.03 (0.76)	.69
	6	297	2.99 (1.44)	.96
	7	297	1.03 (1.19)	.95
	8	297	1.83 (1.25)	.96

9	297	2.41 (1.57)	.97
10	297	1.30 (1.07)	.97
11	297	4.48 (1.39)	.61
12	297	2.63 (1.39)	.48
13	297	2.56 (1.05)	.53
14	297	2.28 (1.00)	.63

Note. Insight indicators are ANAorg (anagrams) administered in an originality condition and SCR (scrabble tasks), either administered in an originality condition (SCRorg) or a fluency condition (SRCflu).

Table SM 4

Fit Indices of the Measurement Models of all Creativity Tasks on the Item Level for Study 1

Model	χ^2 (*df*)	*p*	CFI	RMSEA [90% CI]	SRMR	ω
SA	13.00 (9)	.16	.99	.06 [.00 - .12]	.03	.84
IN	202.71(135)	.00	.91	.06 [.04 - .07]	.06	.90
FF	2.70 (2)	.26	.99	.05 [.00 - .18]	.03	.70
RF	19.59 (9)	.02	.96	.09 [.03 - .15]	.04	.84
CO	59.94 (54)	.37	.97	.02 [.00 - .06]	.06	.64
NI	32.17 (27)	.23	.97	.04 [.00 - .08]	.05	.75

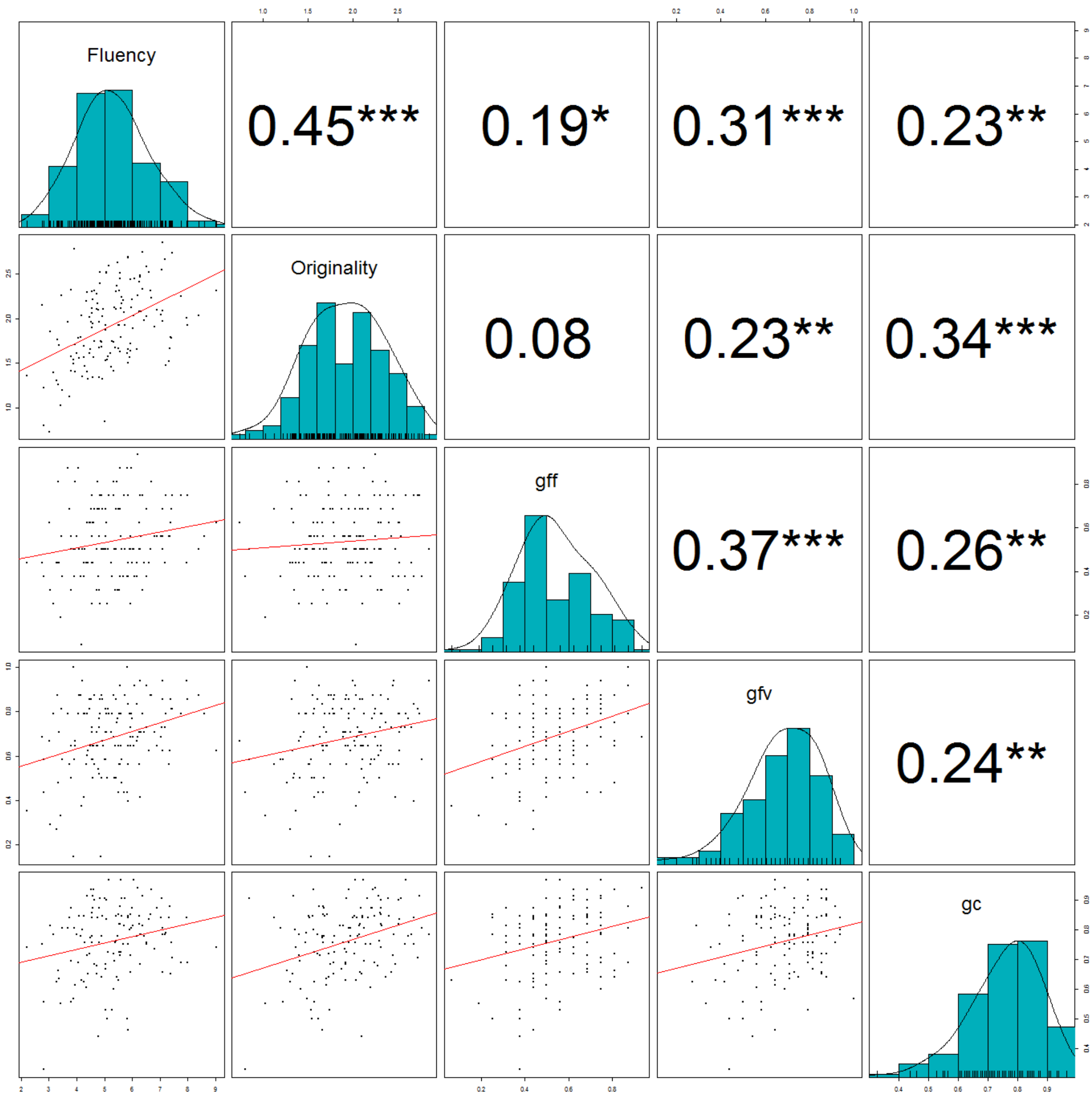
Note. Fluency indicators are SA (similar attributes), IN (inventing names), FF (figural fluency), and RF (retrieval fluency). CO (combining objects) and NI (nicknames) are originality indicators. ω = factor saturation (McDonald, 1999); CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean-Square Residual.

Table SM 5

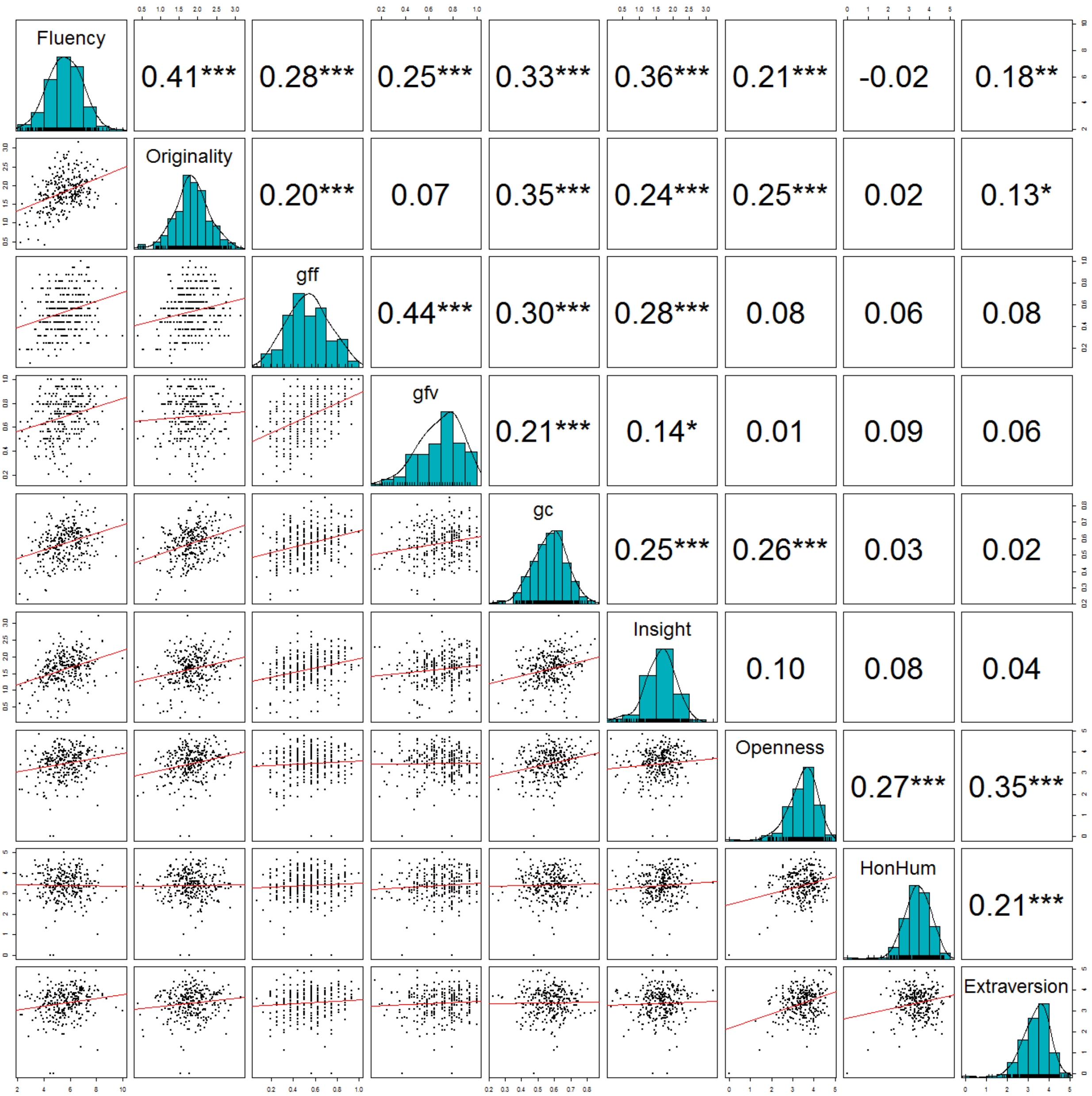
Fit Indices of the Measurement Models of all Creativity Tasks on the Item Level for Study 2

Model	χ^2 (*df*)	*p*	CFI	RMSEA [90% CI]	SRMR	ω
SA	25.94 (9)	.00	.97	.08 [.05 - .12]	.03	.79
IN	224.24 (135)	.00	.95	.05 [.04 - .06]	.04	.91
FF	1.88 (2)	.39	1.00	.00 [.00 - .11]	.01	.65
RF	54.23 (9)	.00	.87	.13 [.10 - .17]	.06	.74
CO	69.16 (54)	.08	.91	.03 [.00 - .05]	.04	.61
NI	40.90 (27)	.04	.98	.04 [.01 - .07]	.03	.81

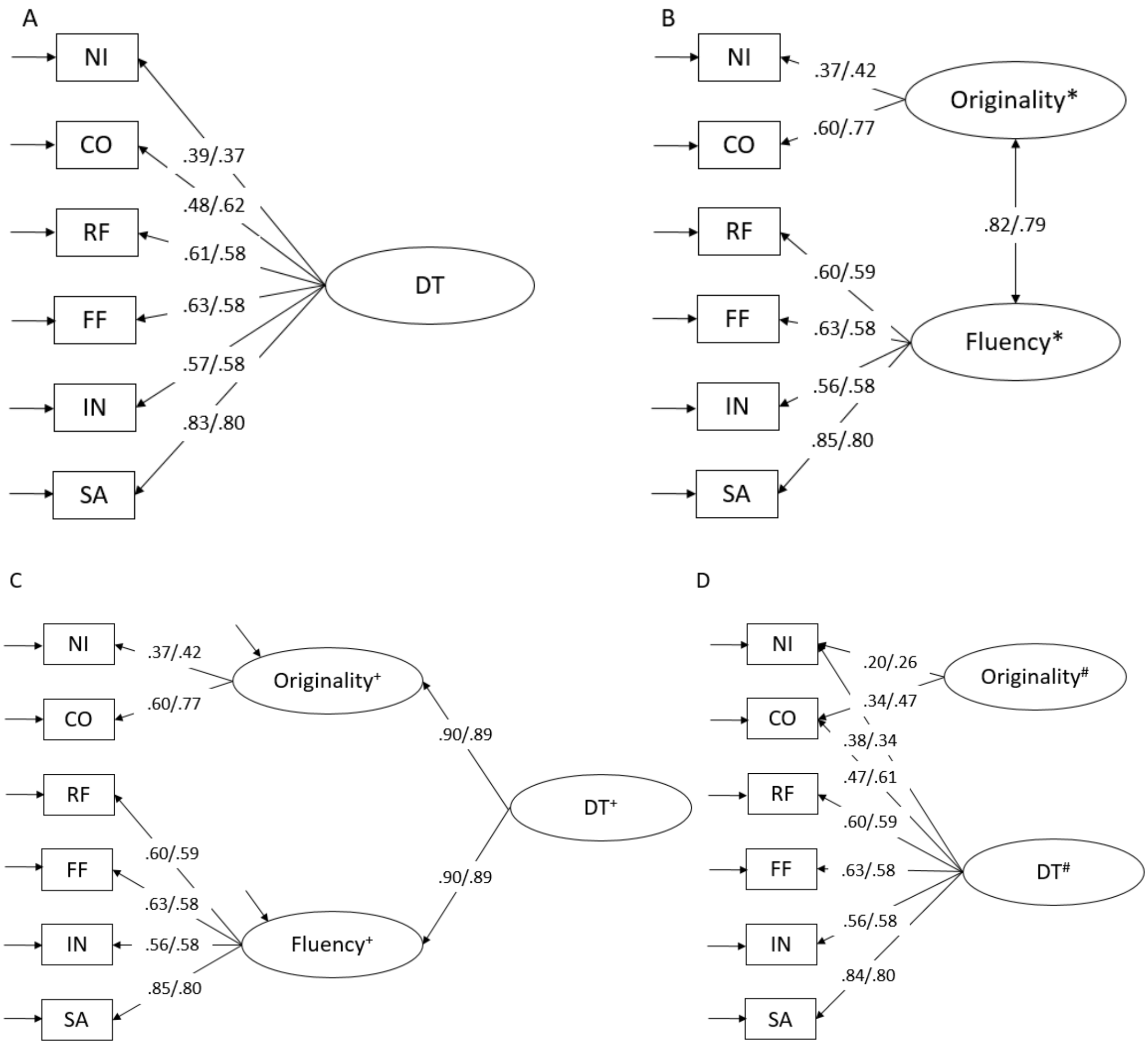
Note. Fluency indicators are SA (similar attributes), IN (inventing names), FF (figural fluency), and RF (retrieval fluency). CO (combining objects) and NI (nicknames) are originality indicators. ω = factor saturation (McDonald, 1999); CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean-Square Residual.



SM Figure 1a. Correlations and bivariate scatterplots for Study 1 between the manifest scores for fluency and originality, fluid intelligence (gff = figural, gfv = verbal) and crystallized intelligence (gc); * < .05; ** < .01; *** < .001.



SM Figure 1b. Correlations and bivariate scatterplots for Study 2 between the manifest scores for fluency and originality, fluid intelligence (gff = figural, gfv = verbal) and crystallized intelligence (gc), insight, and personality (openness, honesty-humility [HonHum] and extraversion); * < .05; ** < .01; *** < .001.



SM Figure 2. Competing measurement models of DT as depicted in Figure 1 in the paper including standardized loadings for Study1/Study2. Indicators are 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. The factor variances of the latent variables were fixed to 1. All factors were scaled using unit variance identification constraints (Kline, 2015).

Table SM 6

Correlations and Regressions between Latent Variables for the Model Depicted in Figure 3 When all Links Are Allowed Except Relations between Superordinate and Associated Subordinate Factors

	Insight	DT[#]	Originality[#]	Openness	Extraversion	Honesty-Humility
Mental Speed (MS[#])	*r* = -.02 (*p* = .821; *CI* = -.16-.13)	*β* = .06 (*p* = .355; *CI* = -.07-.19)	*β* =.14 (*p* = .292; *CI* = -.12-.39)	*r* = -.02 (*p* = .833; *CI* = -.16-.13)	*r* =.03 (*p* = .679; *CI* = -.11-.17)	*r* = -.13 (*p* = .068; *CI* = -.38-.01)
General Intelligence (g)	***r* =.43 (*p* < .001; *CI* = .29-.57)**	***β* =.40 (*p* < .001; *CI* = .27-.54)**	*β* =.02 (*p* = .926; *CI* = -.30-.33)	*r* =.11 (*p* = .167; *CI* = -.05-.26)	***r* =.15 (*p* = .037; *CI* = .00-.29)**	*r* =.10 (*p* = .195; *CI* = -.05-.25)
Crystallized Intelligence (gc[#])	***r* =.21 (*p* = .008; *CI* = .05-.36)**	***β* =.38 (*p* < .001; *CI* = .22-.54)**	*β* =.20 (*p* = .274; *CI* = -.16-.56)	***β* = .36 (*p* < .001; *CI* = .21-.51)**	***r* = -.16 (*p* = .034; *CI* = -.31--.01)**	*r* = -.07 (*p* = .402; *CI* = -.23-.09)
Insight		***r* = .34 (*p* < .001; *CI* = .19-.50)**	*r* = -.13 (*p* = .290; *CI* = -.38-.11)	*r* = .13 (*p* = .096; *CI* = -.02-.29)	*r* = .07 (*p* = .331; *CI* = -.07-.21)	*r* = .10 (*p* = .172; *CI* = -.05-.26)
DT[#]				*β* = .08 (*p* = .414; *CI* = -.11-.27)	***β* = .17 (*p* = .026; *CI* = .02-.33)**	*β* = -.15 (*p* = .055; *CI* = -.30-.00)
Originality[#]				*β* = .05 (*p* = .770; *CI* = -.31-.41)	*β* =.02 (*p* = .876; *CI* = -.28-.33)	*β* = -.06 (*p* = .698; *CI* = -.35-.23)
Openness					***r* = .43 (*p* < .001; *CI* = .30-.55)**	***r* = .36 (*p* < .001; *CI* = .23-.50)**
Extraversion						***r* = .29 (*p* < .001; *CI* = .15-.42)**

Note. $\chi^2(312) = 485.09$, $p < .01$, CFI = .93, RMSEA = .04, SRMR = .05; coefficients in bold type are significant with $p < .05$; CI = 95% Confidence Interval.

Table SM 7

Correlations and Regressions between Latent Variables for the Model Depicted in Figure 3

	Insight	DT[#]	Originality[#]	Openness	Extraversion	Honesty-Humility
Mental Speed (MS[#])	r =-.01 (p = .912; *CI* = .15-.13)	β =.09 (p = .188; *CI* = -.04-.22)				
General Intelligence (g)	**r =.42 (p < .001; *CI* = .28-.56)**	**β =.41 (p < .001; *CI* = .28-.55)**	β =.02 (p = .878; *CI* = -29-.34)			
Crystallized Intelligence (gc[#])	**r =.21 (p = .007; *CI* = .06-.36)**	**β =.38 (p < .001; *CI* = .23-.54)**	β =.16 (p = .372; *CI* = -.19-.50)	**β =.34 (p < .001; *CI* = .19-.49)**		
Insight		**r =.35 (p < .001; *CI* = .20-.51)**	r = -.14 (p = .248; *CI* = -.39-.10)			
DT[#]				β =.07 (p = .429; *CI* = -.11-.25)	**β =.16 (p = .023; *CI* = .02-.31)**	**β = -.16 (p = .024; *CI* = -.30--.02)**
Originality[#]				β =.08 (p = .649; *CI* = -.27-.43)	β =.01 (p = .894; *CI* = -.29-.31)	β = -.07 (p = .619; *CI* = -.36-.22)
Openness					**r =.41 (p < .001; *CI* = .29-.54)**	**r =.36 (p < .001; *CI* = .22-.49)**
Extraversion						**r =.28 (p < .001; *CI* = .15-.42)**

Note. Bold correlations are significant with p < .05; CI = 95% Confidence Interval.

Manuscript IV: A Reappraisal of the Threshold Hypothesis of Creativity and Intelligence

Supplementary Material

Published Online in the Open Science Framework – Center for Open Science:

https://osf.io/6fxv5/

Table S1

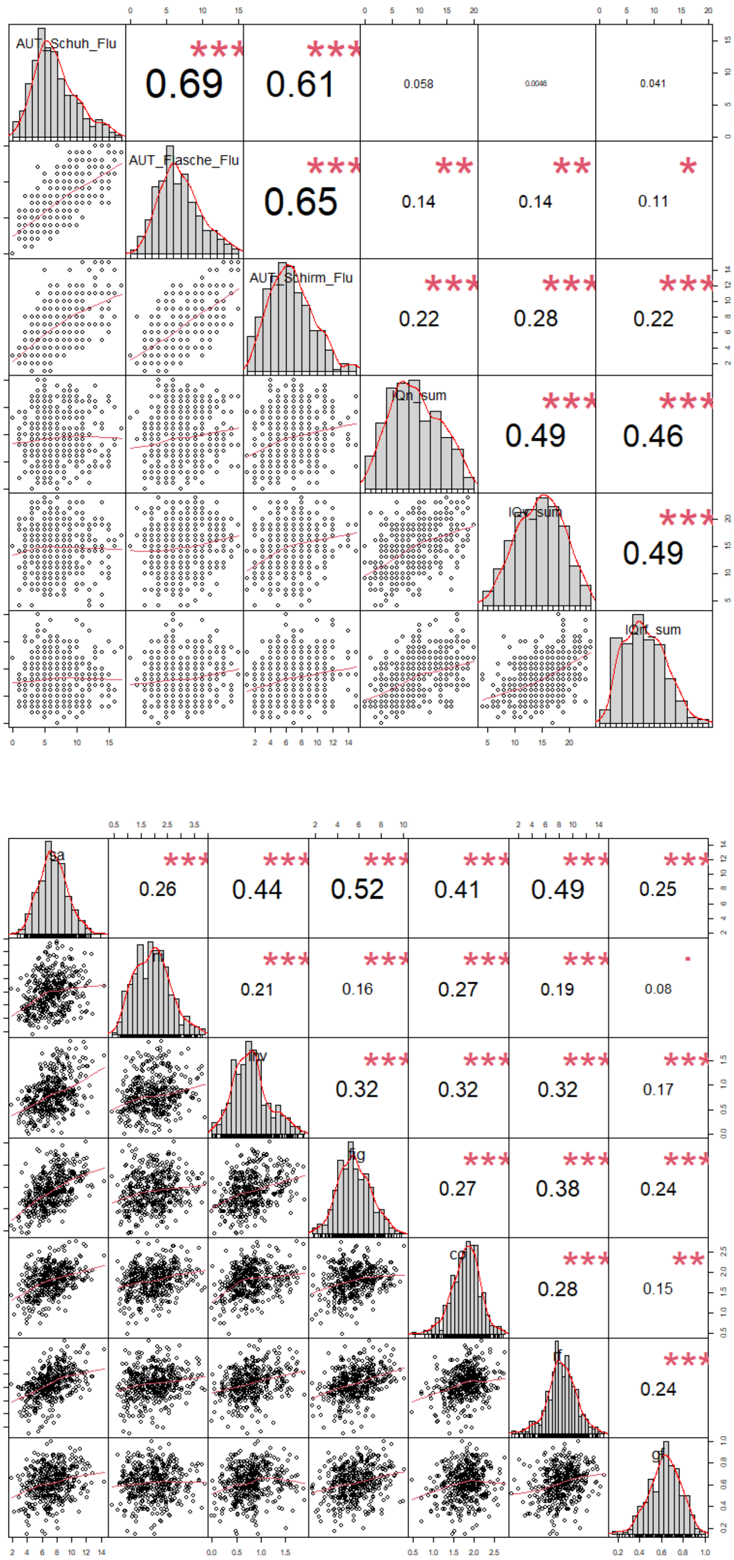
Descriptive Statistics for all Indicators

Dataset	Construct	Task	*N*	*Mean* (*SD*)
1	DT	Shoe	453	6.79 (3.26)
		Bottle	453	7.00 (2.87)
		Umbrella	453	6.71 (2.78)
	IQ	Numerical	451	9.39 (4.55)
		Verbal	454	14.72 (4.28)
		Retrieval	455	8.30 (3.76)
2	DT	SA	421	7.24 (2.07)
		NI	425	1.97 (0.67)
		IN	423	0.80 (0.36)
		FF	427	5.64 (1.59)
		CO	424	1.79 (0.35)
		RF	415	8.51 (2.16)
	IQ	Verbal and Figural	423	0.62 (0.15)

Note. Dataset 2: Fluency indicators are SA (similar attributes), IN (inventing names), FF (figural fluency), and RF (retrieval fluency). CO (combining objects) and NI (nicknames) are indicators of Originality that were only instructed for originality. Fluency indicators were rated regarding the quantity of correct responses, originality was rated regarding the quality of a single given response.

Figure S1

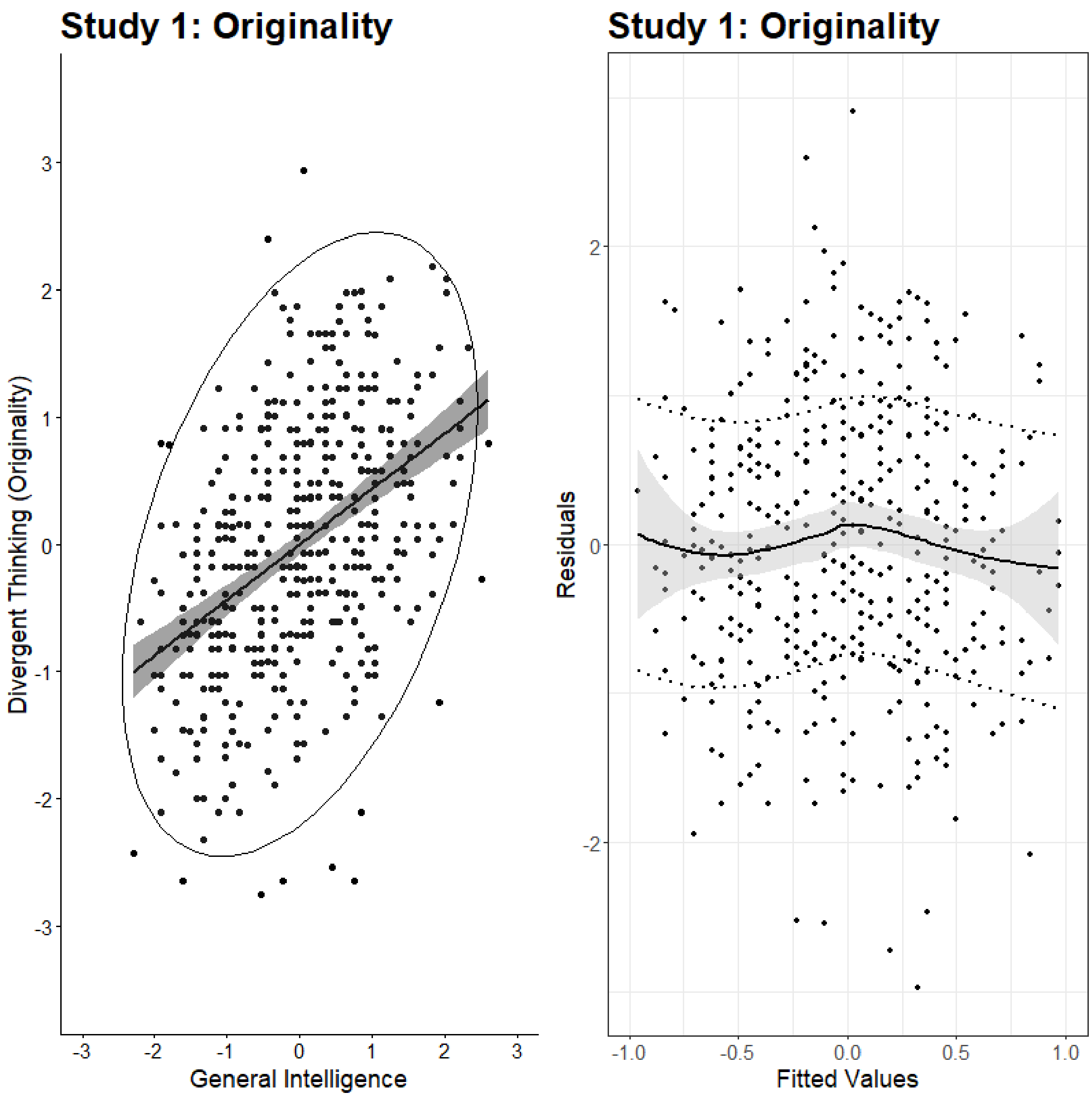
Correlations and Bivariate Scatterplots Between Manifest Scores



Note. Datasets 1 upper figure and 2 lower figure; * < .05; ** < .01; *** < .001.

Figure S2

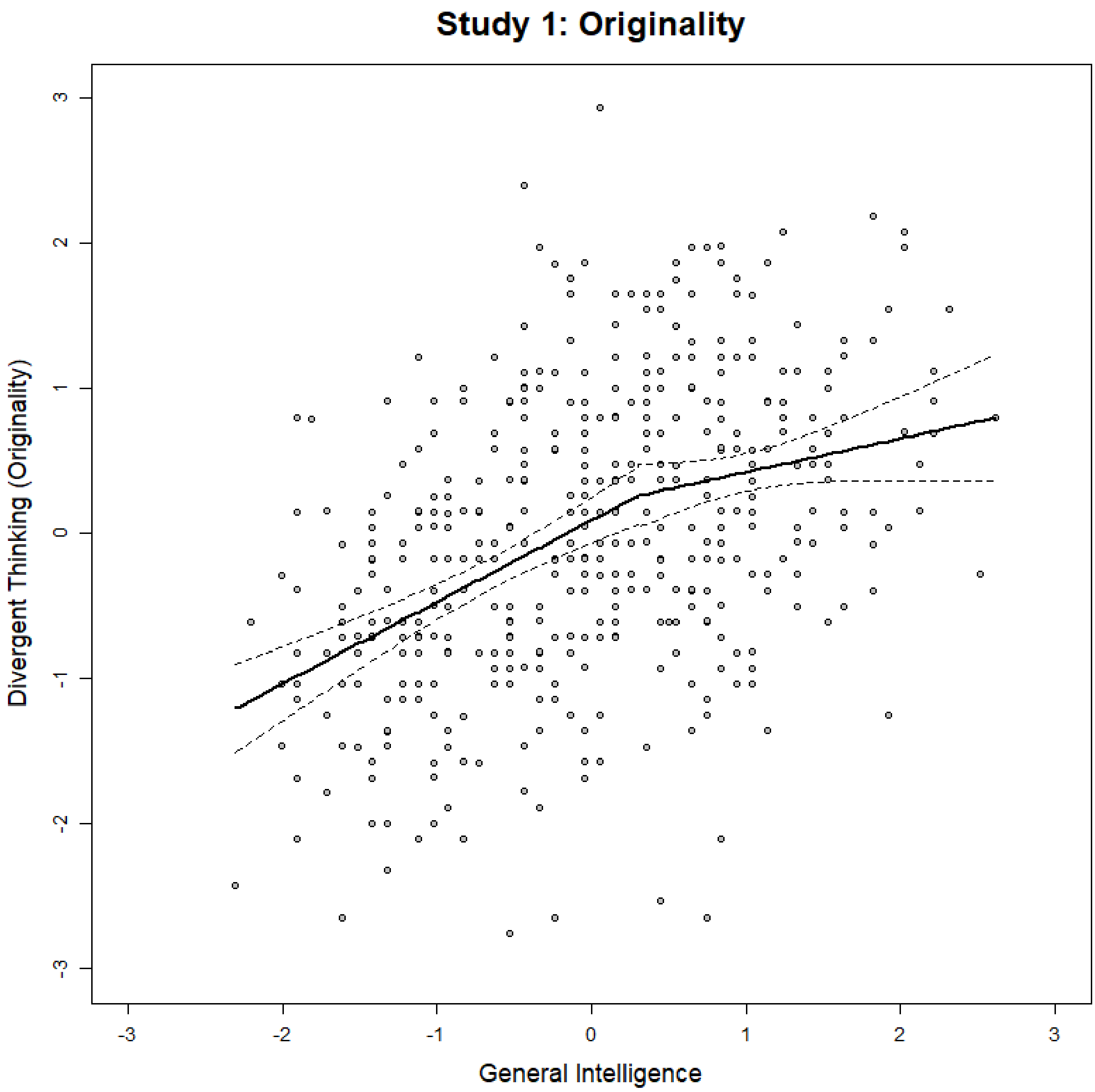
Scatterplot and Heteroscedasticity Plot in Study 1: Originality.



Note. Scatterplot (including a 0.95 confidence interval and an ellipse) for the correlation between divergent thinking and intelligence are presented upper plots. Heteroscedasticity plots (lower plots) include standard errors (grey) and standard deviations of the fitted values (dashed line).

Figure S3

Segmented Regression Analysis in Study 1: Originality.



Note. Breakpoint for the relation between general intelligence and divergent thinking. Dotted line = 95% CI.

Figure S4

Factor Variances at each Focal Point along the Intelligence Continuum in Study 1: Originality.

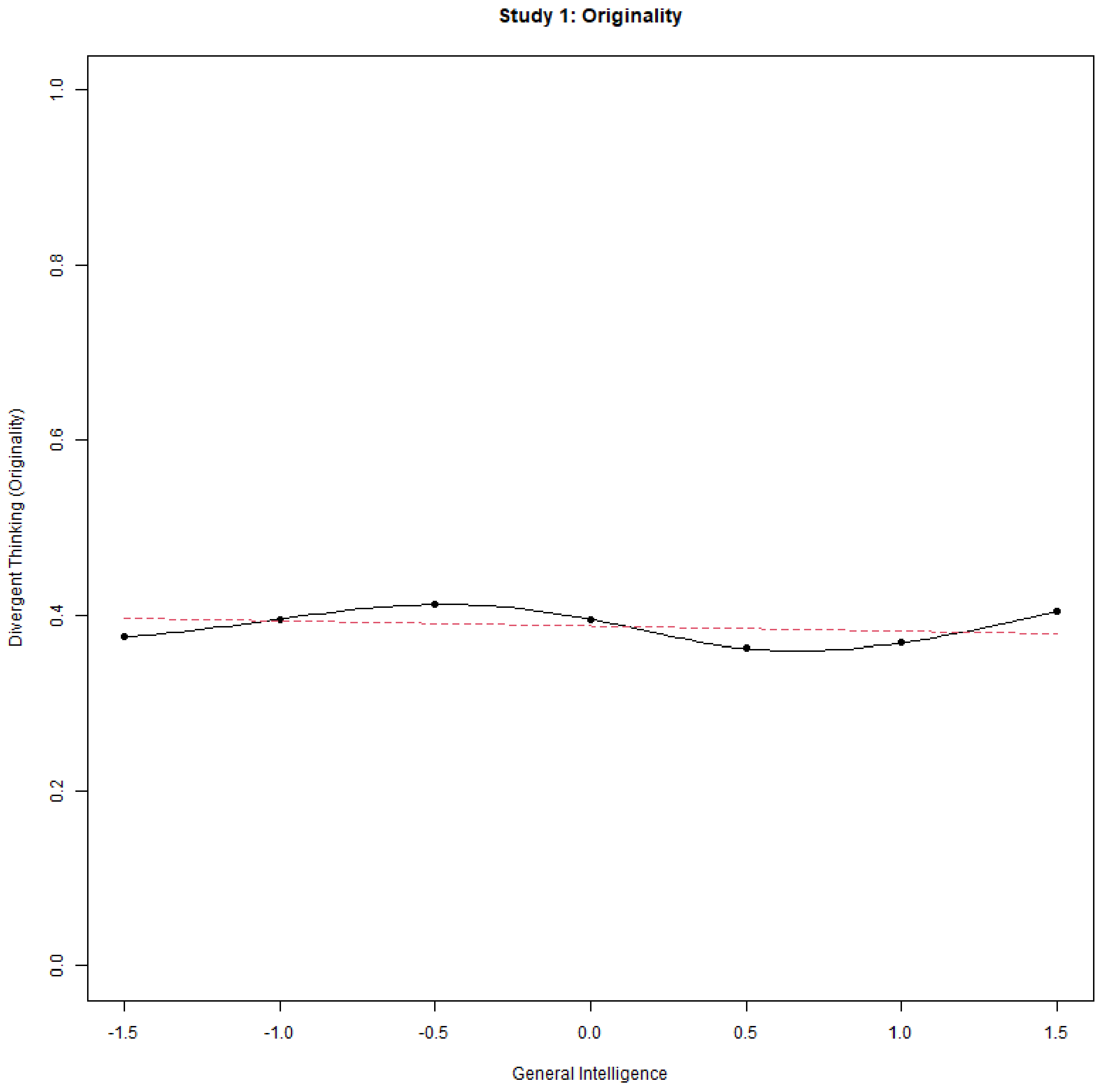
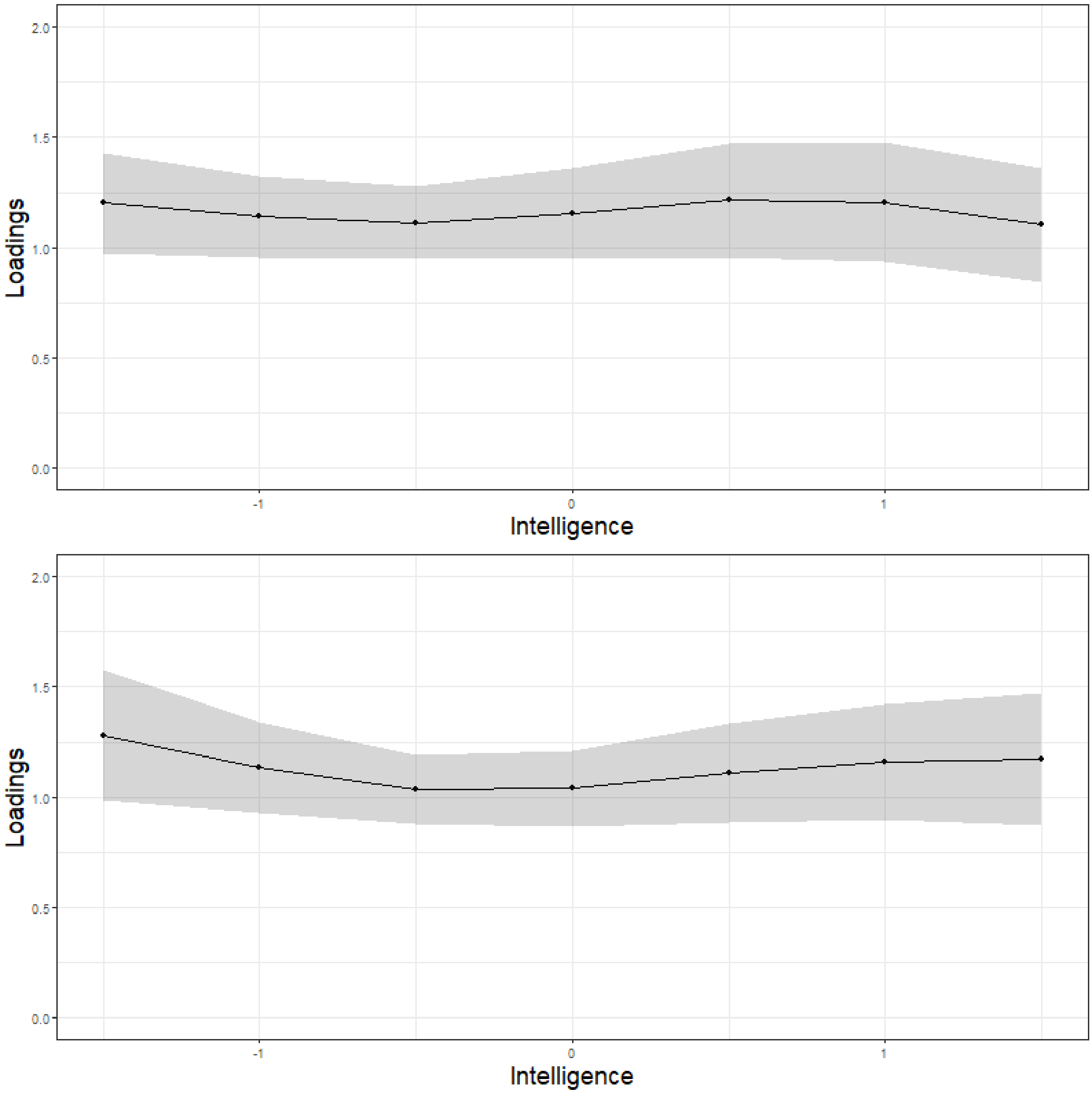


Figure S5

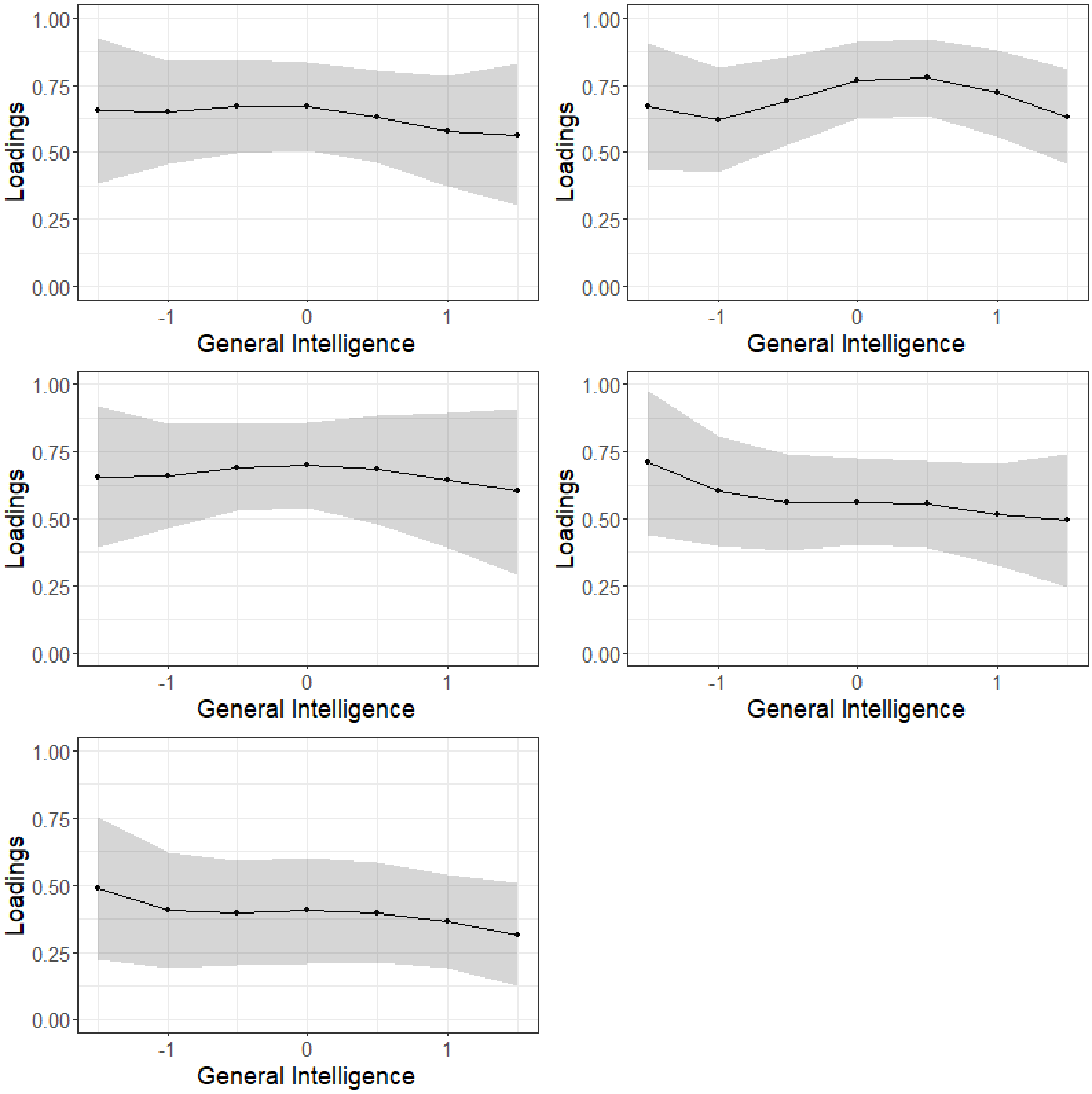
Loadings at the Focal Points in Dataset 1.



Note. Item 2 displayed in the upper figure and item 3 in the lower figure.

Figure S6

Loadings at the Focal Points in Dataset 2.



Note. From upper left: item 2, 3, 4, 5 and 6