Contemporary Educational Psychology 53 (2018) 1–14

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Relational reasoning and divergent thinking: An examination of the threshold hypothesis with quantile regression

Denis Dumas

Morgridge College of Education, Department of Research Methods and Information Science, University of Denver, Denver, CO 80208, United States

ABSTRACT

Relational reasoning (RR) and divergent thinking (DT) are two critical antecedents of creative problem solving, but the relation between them is not currently well understood psychologically, limiting efforts to support these constructs through education. The threshold hypothesis (TH) is currently the dominant explanation for the relation between RR and DT, and posits that RR fundamentally supports DT, but only up to a point. In this study, quantile regression was used to test the TH among RR and two separate dimensions of DT: originality and fluency. Results generally supported the TH in regards to originality, with RR being significantly positively related to originality, but only in students at or below the median of the originality distribution. However, the TH was not upheld for fluency, which was only significantly predicted by RR at the top (i.e. 9th decile) of the fluency distribution. In general, results suggest that direct instructional intervention of RR strategies may be most supportive of creativity for those students who are simultaneously highly fluent but low-original thinkers.

1. Introduction

Perhaps more than at any other time in recent history, the students of today must be prepared for a future in which the geopolitical, environmental, economic, and technological landscape is highly uncertain. Moreover, current predictions concerning the scope of some of these challenges (e.g., climate change; Wilson et al., 2017) forecast that today's young learners will reach adulthood at a point in which major crises have worsened, and solutions to humanity's ongoing problems are critically necessary. Because of the unprecedented nature of such challenges, it is clear that the learners of today must be equipped, as much as possible, with the ability to pose novel and useful solutions to existing problems: a construct termed *creative problem solving* (CPS; Hardy & Gibson, 2017; Orzechowski, Kruchowska, Gruszka, & Szymura, 2017; Parnes, 1961). However, it remains unclear within the educational and psychological research literatures how such a construct develops in students, and how it can best be supported through instruction (Alfonso-Benlliure & Santos, 2016; Silvia, Christensen, & Cotter, 2016; Yi, Hu, Plucker, & McWilliams, 2013).

One principal reason why CPS is not yet well-understood in the literature is because the enactment of CPS appears to be driven by a large number of other constructs including, but not limited to, motivational traits (Hennessey, 2010), emotional states (Hao, Liu, Ku, Hu, & Runco, 2015), prior domain knowledge (Mayer, 2016), and metacognitive strategy use (Hargrove & Nietfeld, 2015). Some researchers have also posited that creativity should be defined as an investment, through a stochastic process that depends on how an idea gains value in the future (Sternberg & Lubart, 1995). However, within the existing literature, there do appear to be at least two constructs that receive perennial attention for their strong predictive relation to CPS. First, the ability to generate multiple possible ideas, responses, or solutions to a given problem (whether or not those solutions are correct), termed *divergent thinking* (DT), is a critical antecedent of productive CPS (Acar & Runco, 2014; An, Song, & Carr, 2016). Second, the ability to discern complex patterns within information, draw connections among ideas, or coordinate multiple points-of-view, termed *relational reasoning* (RR), also appears, based on accumulated empirical evidence, to be necessary for successful CPS (Dumas, Schmidt, & Alexander, 2016; Green, Kraemer, Fugelsang, Gray, & Dunbar, 2012).

However, the current understanding of these two constructs, their relation to each other, and their relation to CPS, is fundamentally undermined by the observation that the relation between RR and DT appears to differ substantially depending on the students within whom the constructs are measured (e.g., Karwowski et al., 2016). For example, the relation between RR and DT can differ markedly depending on students' age, level-of-schooling, or giftedness identification, implying that major differences exist in the cognitive strategies that students of different levels utilize to engage in both RR and DT (e.g., Alban-Metcalfe, 1978; Jauk, Benedek, Dunst, & Neubauer, 2013; Runco & Albert, 1986). Following this pattern is the related observation that even for students of the same age and level of schooling, the relation between DT and RR (or other similar cognitive constructs, such as fluid

E-mail address: denis.dumas@du.edu.

https://doi.org/10.1016/j.cedpsych.2018.01.003

Available online 16 January 2018

intelligence) can differ (Karwowski et al., 2016; Shi, Wang, Yang, Zhang, & Xu, 2017). In particular within the creativity literature, the prevailing view has been that, among students of generally low cognitive ability, the relation among these constructs should be quite strong, while in students of generally high cognitive ability, the relation among these constructs should be weaker (Guilford, 1967; Preckel, Holling, & Wiese, 2006; Runco & Albert, 1986). This line of argument, called the *threshold hypothesis* (TH), remains the dominant explanation of the relation among DT and any cognitive construct (including RR) in the extant literature (e.g., Shi et al., 2017). To date, the TH has been tested many times, with some study findings supporting the hypothesis's predictions (e.g., Jauk et al., 2013), and others not (Kim, 2005; Preckel et al., 2006). However, as will be reviewed in this article, such disagreement in the literature may stem from measurement and methodological inconsistencies within previous studies.

This situation not only adversely affects the field's psychological understanding of creativity, it also implies fundamental educational challenges. For example, if DT and RR are extremely strongly positively related in some groups of students, but very weakly positively or even negatively related in others, future intervention efforts would need to be adjusted accordingly in order to account for the necessary heterogeneous effects of DT or RR instruction. Moreover, if a major goal of education remains the cultivation of CPS, and both DT and RR appear to support CPS but do not relate to each other in a predictable way, then no fully mechanistic understanding of how CPS occurs through both divergent and relational processes is possible. In short, without a reasonable consensus as to how DT and RR relate to one another (and for which students that relation can be found, and why), the field cannot begin to pose and test hypotheses pertaining to how these two constructs interact to produce CPS, and how best to teach them to students.

A recent robust meta-analysis (i.e., Gajda, Karwowski, & Beghetto, 2016) has found a positive correlation ($r = 0.22$ on average) between creativity and academic achievement, drawing attention to the fact that students with greater CPS ability tend to perform better than their lower CPS peers on a variety of academic outcomes. However, unlike research on the TH, the literature pertaining to the relation between creative attributes and academic achievements or learning outcomes rarely if ever consider the curvilinear relation posited by the TH (Gajda et al., 2016), illustrating that the current educational understanding of creative abilities such as DT, and their effect on other attributes related to learning, such as RR, may be incomplete.

Given this state-of-affairs, the current study aims to update the measurement tools, as well as statistical procedures, used to test the threshold hypothesis in order to much more fully understand the relation between DT and RR across students of diverse ability levels. Before introducing the current study, the existing literature on DT and RR must be thoroughly reviewed, with a particular emphasis on aspects of the extant literature (e.g., measurement) that may affect the interpretation of previous study findings. Then, prior findings related to the threshold hypothesis, as well as those methodological procedures that allows the TH to be tested, will be reviewed as exhaustively as possible. Next, given the identified shortcomings of previous work, the goals of the current study will be described before introducing the current study methodology and results.

1.1. Divergent thinking

As opposed to reasoning processes, perhaps more commonly studied in the literature, that require individuals to "converge" on a particular correct solution to a given problem, DT is a process in which students "diverge" their thinking by producing a number of possible ideas (Hudson, 1968; Lewis & Lovatt, 2013). Further, as a way to distinguish DT from actual creative achievement in a given domain of expertise, it is sometimes referred to in the existing literature as "creative potential" (e.g., Acar & Runco, 2014). From an educational psychology perspective, DT may be better described as a domain-general creative ability, whose measurement is designed (similarly to many reasoning or intelligence measures) to be as free from the influence of prior domain knowledge as possible in order to tap individuals' relatively pure ideational ability. Of course, individual responses on any DT task are necessarily constrained by prior knowledge and experience (Hass, 2017), but by relying on stimuli with which most students have a similar level of knowledge, DT tasks attempt to minimize the effect of prior knowledge on the score distribution.

For example, by far the most commonly utilized existing measure of DT is the Alternate Uses Task (Plucker & Makel, 2010; Torrance, 1972), which requires participants to generate as many novel uses of a given every-day object as they can. In this way, data from DT tasks such as the AUT are typically interpreted as evidence of multiple dimensions of DT including the quantity of ideas generated (i.e., *ideational fluency*; Runco et al., 2011), and the relative novelty of those ideas (*originality*; Hass, 2017). It should be noted that some researchers have sought to quantify yet more dimensions of DT, including flexibility (Acar & Runco, 2017), and elaboration (Mohr, Sell, & Lindsay, 2016). However, efforts to quantify these dimensions reliably have not entirely come to fruition, and as such many researchers today focus on the two dimensions of DT for which strong psychometric evidence (e.g., Dumas & Dunbar, 2014) exists: fluency and originality.

In part because fluency is very straight-forward to operationalize by simply counting the number of uses generated by participants on the AUT, it is without doubt the most commonly measured dimension of DT (e.g., Turner, 1999). Indeed, it is not uncommon in the extant literature to use terms such as "fluency" and "DT" essentially as synonyms (e.g., Jones & Estes, 2015). Also related to its straight-forward calculation, fluency scored DT measures tend to exhibit good reliability, because participant ability to generate a number of ideas tends to be relatively stable across AUT prompts (e.g., book, brick; Karwowski et al., 2016). Perhaps in part due to this typically high reliability, fluency scores have proven useful in the creativity literature for decades, regularly predicting relevant CPS outcomes in a number of domains of expertise, and even longitudinally across time (Runco, Millar, Acar, & Cramond, 2010). Such predictive power has led to the claim, often repeated, that DT measures have higher validity coefficients than do cognitive assessments (Plucker & Makel, 2010). However, despite this documented usefulness of fluency scores, such scores provide no information about the novel quality of the ideas generated by an individual. For this, originality scoring is needed.

In contrast to fluency, which has had a straight-forward and agreed-upon scoring procedure for decades, the quantification of the originality of participant-generated ideas has been anything but simple. As an illustration, if students were prompted on the AUT to generate uses for a toothbrush, those uses may range from the quotidian (e.g., brush your teeth) to the relatively original (e.g., replacement tent-stake). But, *how much* more original is the latter response than the former? Typically during the 20th century, this question was answered by tasking panels of individuals (usually drawn from the same population as the study sample) to score each participant's responses on an originality scale before averaging those originality ratings across the panelists (Sternberg, 2006a). Sometimes, participants themselves have also been asked to rate the originality of their own responses (Silvia, 2011). Unfortunately, because individual raters on any given panel may hold differing views about what constitutes originality, such scoring procedures typically led to relatively low score reliability even when raters were explicitly trained (Sternberg, 2006b). Perhaps more problematic, even when scores were shown to be reliable, the fundamental subjectivity of these scoring procedures undercut the generalizability of conclusions that could be made from them. At times, counts of the number of ideas an individual produces on the AUT (i.e. fluency) are amended to reflect only those responses that are unique in the given sample, such that an original response is defined as a unique response. However, such uniqueness scoring procedures have a number of practical problems such as the confounding of fluency and originality, and the observation that as sample-size increases, the likelihood that any AUT response is unique in the sample, no matter how original that response may be, drastically decreases (Silvia et al., 2008). Therefore, over the last ten-years, researchers have begun to operationalize originality more objectively as the semantic distance between an AUT prompt and a given participant response.

Beginning in the 1990s and becoming more wide-spread in the last decade, modern computational and statistical technology have made it possible, through a technique called *latent semantic analysis*, to objectively quantify the semantic relations among terms (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Dumas & Dunbar, 2014; Landauer, Foltz, & Laham, 1998). LSA is a latent variable statistical technique for quantifying the semantic relation between words or phrases through the analysis of a very large body of text, called a corpus (Beaty, Christensen, Benedek, Silvia, & Schacter, 2017; Green, Kraemer, Fugelsang, Gray, & Dunbar, 2010). In an LSA analysis, word-co-occurrence in such a corpus is represented within a multi-dimensional matrix, and, within this multivariate semantic space, particular word- or phrase-vectors are identified. In order to determine how semantically related a particular pair of words are, the cosine of the angle between their vectors is calculated, yielding a latent-semantic similarity statistic that, in the same way as a Pearson correlation, falls between −1 and 1. If a participant's response (e.g., brush your teeth) is strongly positively associated with the AUT prompt (e.g., toothbrush), this is an indicator that the response is not very original. In contrast, responses that are more original (e.g., replacement tent-stake) yield a weaker similarity statistic, and perhaps even a negative similarity statistic with the original prompt. It may be interesting to note that one pioneer of the application of LSA was educational psychologist Walter Kintsch (2000), who used the method to inform studies of text and metaphor comprehension, and today the method is required for automated essay scoring algorithms (Foltz, Streeter, Lochbaum, & Landauer, 2013). Previous measurement work (Dumas & Dunbar, 2014; Forster & Dunbar, 2009) has suggested that LSA may yield reliable and objective measurements of originality, and that fluency and originality can best be described as discrete but positively linearly related constructs, with those participants who generate more responses to the AUT also providing more original responses on average. Moreover, LSA is rapidly becoming the measurement standard for quantifying originality among creativity researchers (e.g., Hass, 2017), and as such, it will be utilized in the current study to better understand the relations among DT dimensions and RR.

1.2. Relational reasoning

Because RR is broadly conceptualized as the ability to ascertain patterns within any stream of information (Alexander, 2012; Dumas, Alexander, & Grossnickle, 2013), the construct can be conceptualized as multidimensional, encompassing at least four associated relational processes. Specifically, if higher-order relations of similarity are discerned among multiple concepts, an *analogy* can be identified (Holyoak, 2012; Richland, Begolli, Simms, Frausel, & Lyons, 2016). For example, one commonly drawn analogy in science education is between an atom and the solar system, which share some structural features (Coll, 2008). In contrast, an *anomaly* occurs when a relation of discrepancy between an expected and observed pattern is detected (Filik & Leuthold, 2008). For example, anomalous findings are frequently implicated as critical in the conceptual change literature, in which students must observe their previous concepts to be demonstrably false in order to learn a new, more accurate concept (e.g., Chinn & Malhotra, 2002). Further, if two or more ideas or concepts are determined to be incompatible or paradoxically related, an *antinomy* has been identified (Resnick, Davatzes, Newcombe, & Shipley, 2016). For instance, in medical education, students are taught to determine whether a given patient's symptoms are incompatible with a given pathology, and therefore rule-out that diagnosis (Dumas, Alexander, Baker, Jablansky, & Dunbar, 2014). Finally, ideas are *antithetical* if they are oppositionally related to one another along a single dimension or scale (Kendeou, Butterfuss, Van Boekel, & O'Brien, 2016). Such a relation is frequently built-into refutation texts, which systematically repudiate mis-conceptions commonly held by students, typically in science (Danielson & Sinatra, 2016).

Taken together, these four forms of RR have been empirically shown to support student learning across the gamut of academic domains and levels-of-schooling, from early elementary reading instruction (Ehri, Satlow, & Gaskins, 2009), to middle-school mathematics (Richland et al., 2016), high-school chemistry (Trey & Khan, 2008), undergraduate meteorology (Braasch & Goldman, 2010), and graduate level engineering (Dumas et al., 2016), among many others. RR's general influence on learning and academic achievement is perhaps not surprising, because RR as a construct is conceptually similar to intelligence, which was originally defined as the "cognition of relations" among early intelligence researchers (Spearman, 1927; p. 165). Given this conceptual similarity, it also follows that the most commonly utilized measures of RR within the literature are also associated with the assessment of intelligence: visuo-spatial matrix analogies such as those on Raven's Progressive Matrices (RPM; e.g., Baldo, Bunge, Wilson, & Dronkers, 2010), and four-term verbal analogies of the form A:B::C:D (Green et al., 2010).

However, more recently, researchers have begun to construct measures of RR that more meaningfully incorporate the multi-dimensional nature of the construct, and are based on a more precise understanding of the cognitive processes required to solve relational problems. For example, the Test of Relational Reasoning (TORR; Alexander, Dumas, Grossnickle, List, & Firetto, 2015; Dumas & Alexander, 2016) is a visuo-spatial measure with four scales, one for each of the forms of RR (i.e., analogy, anomaly, antinomy, and antithesis). To date, empirical evidence has suggested that TORR scores are highly reliable, and capable of predicting a variety of academic variables, including CPS outcomes within the domain of engineering (Dumas et al., 2016). Moreover, because the TORR has been calibrated and normed using multidimensional item-response theory methods, its scoring produces a single normative metric, termed the relational reasoning quotient (RRQ), which takes into account the multiple-scale structure of the TORR. For these reasons, the TORR may be the best available measure to utilize when seeking to understand the relation between RR and DT, as is the goal of this study. Therefore, here, RR will be assessed using the TORR.

1.3. Threshold hypothesis

Uncovering the relation between creative abilities, such as the originality and fluency dimensions of DT, and complex cognitive abilities, such as RR, has been a classic problem in the literature, examined over at least the last 50 years (e.g. Torrance, 1962). However, scholarly debate over the precise nature of this relation—and explanation of why that relation occurs—continues to this day, to the effect that no true consensus currently exists in the literature (Kaufman & Plucker, 2011; Silvia, 2015). Early on in the history of creativity research Guilford (1967) forwarded a mechanistic conceptualization of cognitive abilities (bundled together under the term intelligence) as supportive of creativity but only up to a particular point. Specifically, Guilford (1967), posited the hypothesis that intelligence and creativity would be strongly positively correlated among individuals with average or below-average intelligence, while in those with above average intelligence (traditionally thought of as an IQ score above 120; Cho, Nijenhuis, van Vianen, Kim, & Lee, 2010; Kim, 2005), creativity and intelligence would correlate essentially at zero. As previously stated, this theory is termed the *threshold hypothesis* (TH) because it posits a point on the distribution of intelligence after which creativity and intelligence are no longer related. Such a relation is inherently nuanced, because it requires the relation between constructs to differ for different individuals for different mechanistic reasons—something that has proven difficult to quantitatively model in the existing literature (Karwowski & Gralewski, 2013).

One critical caveat to note here about the existing literature on the threshold hypothesis, is that nearly all extant studies of the TH explicitly examine the relation between creative abilities or achievement (e.g., divergent thinking) and a generalized cognitive attribute termed *intelligence* (e.g., Preckel et al., 2006). However, the current study more specifically operationalizes the cognitive ability being measured and modeled as *relational reasoning*. Perhaps interestingly, these two constructs do share some definitional similarity, stemming from early intelligence work from Spearman (1927) and Raven (1941), who broadly conceptualized intelligence as the "cognition of relations" (Spearman, 1927; p. 166). Indeed, the measure that Raven created to tap fluid intelligence (i.e., RPM; 1941) is now more specifically described as a measure of visuo-spatial analogical reasoning, which is considered a form of RR (Baldo et al., 2010).

Focusing on relational reasoning, as opposed to intelligence, holds a number of advantages for the current study. For one, it may help to link empirical research on creativity and DT with the growing literature on RR: two research literatures that are rarely linked, although in the small number of existing cases that they have been, the results have been fruitful (e.g., Dumas et al., 2016; Green et al., 2012). Additionally, the construct of RR has been more explicitly formulated in terms of the cognitive strategies that are required for its enactment then has intelligence, opening the door for direct instruction of RR strategies to students (Grossnickle, Dumas, Alexander, & Baggetta, 2016), and implying that relational reasoning may be a malleable or trainable set of cognitive strategies and skills that educators may leverage, at least in domain-specific instruction (Resnick et al., 2016). Therefore, the results of this study may be able to inform future instructional efforts, and possibly form the basis for future hypotheses related to the effect of RR training outcomes.

Another inherent complication of Guilford's TH is that it pertained to two extraordinarily multidimensional and generalized constructs (i.e., intelligence and creativity) that, as the field progressed, were deconstructed into various components, some of which (e.g., fluid and crystallized intelligence; creative achievement and DT) were measured very differently from one another. Unsurprisingly, very different results related to the TH have been found depending on the measures utilized, and whether those measures were nested within a particular domain of learning (i.e., crystallized intelligence/knowledge measures; creative achievement checklists) or whether they are designed to rely as-solely-as-possible on fluid processing (e.g., RPM; DT tasks). For example, support for the TH has been uncovered when using checklist-style creative achievement questionnaires and knowledge-based crystallized intelligence measures (e.g., Sligh, Conners, & Roskos-Ewoldsen, 2005). However, one recent comprehensive study found no evidence for the TH when utilizing a visuo-spatial measure of RR (i.e., RPM) and a non-verbal, drawing-based divergent thinking task (Karwowski et al., 2016). Moreover, there is evidence that procedural choices in certain studies, such as the choice of whether to limit the time in which participants may respond to a measure, may also substantially affect the decision to reject or uphold the TH, because participant speed-of-processing may inflate the correlation among time-limited measures (Rindermann & Neubauer, 2004). Therefore, it is clear that when examining the TH, the particular cognitive and creative constructs chosen, as well the specific measures used to operationalize those constructs, and even the procedures used to gather data from participants (not to mention the characteristics of the participants themselves), all deeply affect the findings.

Besides these measurement and procedural concerns, decisions about whether or not to support or reject the TH are also deeply affected by the statistical methods used to run those tests. For many years in the TH literature, the most commonly utilized methodology was simply to compute a bivariate correlation between two scores below a particular cognitive threshold (typically IQ = 120) as well as a bivariate correlation above that threshold, and then compare the significance of those correlations (e.g. Cho et al., 2010; Runco & Albert, 1986). However, such a procedure holds a number of serious shortcomings including the necessity to a priori designate an arbitrary threshold point, and the strong likelihood that range-restrictions on the measured variables above the threshold will artificially lower the correlation, therefore leading to an erroneous support of the TH (Karwowski & Gralewski, 2013; Sligh et al., 2005). In more recent years, more advanced methods have been applied such as multi-group path models, in which the relation between two variables is quantified across multiple groups of students, with group-membership typically defined by their cognitive scores being above or below a particular threshold (e.g., gifted and non-gifted students; Preckel et al., 2006). Although this multi-group procedure clearly leads to more meaningful results than does a comparison of bivariate correlations, it does not necessarily solve all earlier methodological problems however, because it also relies on an a priori designated, fundamentally arbitrary, threshold point to determine group membership.

In order to move past these limitations, researchers have recently begun to rely on two other methodologies for testing the TH: segmented regression (Jauk et al., 2013; Shi et al., 2017) and necessary-condition-analysis (NCA; Karwowski et al., 2016). Because segmented regression does not require an a priori designated threshold point, it can be utilized in an exploratory way to estimate the point at which cognitive abilities cease to predict creative abilities (Muggeo, 2003). Utilizing this methodology, a threshold on the relation between ideational fluency and grammatical reasoning ability has been detected lower than one standard deviation below the mean of grammatical reasoning (i.e., IQ = 81) and, for the relation between creative activity and visuo-spatial ability, as high as IQ = 121 (Karwowski et al., 2016). Although, in other studies that measured intelligence with an RR task (i.e., RPM; Shi et al., 2017), the threshold was found to be close to the mean of RR (i.e., IQ = 109) for fluency, and higher (i.e., IQ = 117) for originality. Based on these findings from segmented regression studies, it is clear that Guilford's (1967) original hypothesis of a threshold at IQ = 120 is not sufficient to explain the relation between all cognitive and creative abilities. Like segmented regression, NCA is another popular modern method for testing the TH that does not require the a priori specification of an arbitrary threshold (Karwowski et al., 2016). In essence, NCA tests the hypothesis that, in a given dataset, participants manifest both high creative and high cognitive abilities, both low creative and low cognitive abilities, and low creative and high cognitive abilities, but *not* high creative and low cognitive abilities. If such a pattern is observed in the data, NCA produces inferential statistics that support the hypothesis that one variable (cognitive ability) is a necessary-but-not sufficient condition for the other (creative ability). To date, previous NCA analyses with an RR measure (i.e., RPM) as well as DT tasks have revealed a necessary-but-not-sufficient pattern of moderate strength (Karwowski et al., 2016; Shi et al., 2017), implying that RR may be indeed necessary, although not sufficient, for DT. In addition, NCA has been recently utilized to examine the relation between RPM scores in childhood and adult creative accomplish, revealing a further necessary-but-not-sufficient relation in the literature (Karwowski, Kaufman, Lebuza, Szumski, & Firkowska-Mankiewicz, 2017).

Although there is no doubt that this current landscape of the TH literature is one that provides meaningful information about the relation between these constructs, I contend that current findings are fundamentally limited by an assumption about the TH that has been in place since Guilford's (1967) time: the directionality of the relation between cognition and creativity. Specifically, the TH has always been formulated as pertaining to a threshold point on the distribution of cognitive ability (i.e., IQ = 120). In the past, when bivariate correlations were the main methodological tool, the theoretical directionality of the predictive relation above and below this threshold did not matter, because bivariate correlations are not directional. However, as methods have advanced to include directional predictive methods such as segmented regression and NCA, the theoretical direction of the predictive relation became fundamentally important. As far as I am aware, to date, all studies of the TH using directional methods have solely designated cognitive measures as independent variables (i.e., predictors) and creative variables as dependent variables (i.e. outcomes; Jauk et al., 2013; Karwowski et al., 2016; Shi et al., 2017).

This decision appears to be based on traditional practice within the intelligence research literature, in which fundamental cognitive abilities such as RR or fluid intelligence are typically believed to support, and therefore precede, nearly all other mental abilities including creative abilities (Furnham, 2016; Guilford, 1967). Such a belief in the intelligence research tradition was formalized early-on by Spearman (1927), in his *law of diminishing returns*, in which general intelligence *g* was considered to undergird all other human mental activities, but the influence of *g* on other abilities was hypothesized to weaken as individuals increased in skill on those other abilities. In this way, before Guilford, Spearman actually hypothesized a threshold in the relation between *g* and all other types of mental abilities. Crucially, because Spearman generally considered *g* to be an innate mental capacity, it would probably not have occurred to him to consider the opposite predictive relation, in which other mental characteristics (e.g., DT) supported *g* to various extents. Forty years later, when Guilford (1967) was forwarding the TH, he followed Spearman's (1927) traditional approach by hypothesizing a threshold on the distribution of cognitive ability, such that cognition was conceptualized essentially as a necessary-but-not-sufficient condition for creativity, but the influence of creative abilities (e.g., DT) on cognitive abilities (e.g., *g*, RR) was not of principal interest.

Today, 90 years after Spearman (1927) and 50 years after Guilford (1967), such a traditional conceptualization of these constructs and the relation between them has been called into question, and therefore should not be treated as dogma. For example, a large amount of evidence accumulated over multiple decades has dismantled the belief that cognitive measures are capable of capturing an innate, non-malleable, intellectual capacity (e.g., Feuerstein, Rand, & Hoffman, 1979; McNeish & Dumas, 2017; Sternberg et al., 2002). Moreover, student ability to respond to DT tasks has also been shown to be highly malleable and subject to improvement through some of the same methods used to improve cognitive scores (e.g., Dumas & Dunbar, 2016; Kassim, Nicholas, & Ng, 2014; van de Kamp, Admiraal, van Drie, & Rijlaarsdam, 2015). Additionally, although the evidence remains clear that cognitive abilities do support creative abilities (Jauk et al., 2013; Karwowski et al., 2016), there is also compelling evidence that creative abilities such as DT support students' ability to solve complex cognitive problems (Green et al., 2017). Of course, other individual differences, such as personality (e.g., Silvia, 2008), can also deeply affect both cognitive and creative abilities, as well as the relation between them, complicating the picture. One modern line-of-thinking is that, as it turns out, intelligence and creativity are "pretty similar after all" (Silvia, 2015), because both constructs rely on overlapping executive functions, cognitive and metacognitive strategy use, and particular personality factors. In this way, it appears unlikely that any particular measurable cognitive ability, whether it be *g*, RR, or crystalized knowledge, actually precedes all creative abilities, including fluency and originality, in the way that early intelligence researchers argued. Therefore, I would contend that, in order to more deeply understand the relation between RR and DT, the TH must not only be tested in a scenario in which RR predicts DT—and the threshold point is on the distribution of RR—but must also be tested as DT predicts RR, and the threshold point is on the distribution of DT (i.e., fluency and originality).

This argument also logically arises from a psychological interest in education, in which the value of instructional intervention on one particular variable (e.g., RR) may be built on the expected student gains on a number of other associated constructs (e.g., CPS or DT). However, it is clear in educational psychology that, during instruction, all students do not begin with the same level of ability, grow the same amount in that ability, or benefit the same amount in terms of other associated measured skills (Choi, Elicker, Christ, & Dobbs-Oates, 2016; McNeish & Dumas, 2017). Such a phenomena may be partly explained by differences in the strength of the relations among educationally-relevant psychological constructs across a diverse (in terms of ability level) population of students. Without specific quantitative estimates of these heterogeneous relations, there can be no way to predict how even homogenous learning gains associated with instruction on one ability will lead to heterogeneous improvement on other educationally relevant outcomes. So, from the educational psychology perspective, specific estimates of the relation between constructs (e.g., RR and DT) across the distribution of both variables, are meaningful and of interest. Further, the previously held research paradigm in which cognitive abilities (e.g., intelligence, RR) are assumed to precede creative abilities (e.g., DT) likely does not entirely match reality, and therefore the predictive relations from RR to DT, and vice versa, may be pertinent.

1.4. Goals of current study

In order to address these observed limitations in the existing literature, the current study turns to a methodology that has only rarely been used in education research: quantile regression (Koenker, 2005). Quantile regression allows the quantitative relation among an outcome variable (Y) and one or more predictors (X) to be modeled at multiple researcher-specified points on the distribution of Y. In general, quantile regression is an effective method to model the predictive relations among variables if those relations are hypothesized to differ depending on the level of the one of the variables, as they are in the TH. Perhaps notably, quantile regression is commonly utilized in the economics literature, in which certain outcome variables (e.g., housing prices) relate to various predictors (e.g., house and buyer characteristics) differently depending on the level of the outcome variable (e.g., Zietz, Zietz, & Sirmans, 2008). Within the educational psychology literature, quantile regression was recently applied to understanding the differential effect of various cognitive predictors (e.g., working memory) on reading comprehension for students who differ on their comprehension ability (Language and Reading Research Consortium & Logan, 2017). In the context of the TH, quantile regression may improve upon existing methods because it neither requires an a priori specified threshold nor models the location of a single threshold—instead, quantile regression provides detailed information about slope heterogeneity across the full distribution of the outcome variable. Therefore, quantile regression may be used, in a way that previous methods such as segmented regression and NCA cannot, to quantify the relation between RR and DT (and vice versa) at any point along the distribution of either variable. So, much finer-grain information may be gleaned from a quantile regression analysis than any other previously used method.

Further, as previously implied, the current study seeks to expand upon existing work by examining the TH in both theoretically plausible directions: from RR to DT and from DT to RR. In doing so, it is expected that a much fuller picture of the underlying relationship between these constructs will emerge, and potentially allow for richer educationally-relevant inferences to be made. Although these general study goals are complex, they can be expressed in terms of a number of specific empirical research questions:

1. What are the bivariate correlations among ideational fluency, originality, RR, and each of its forms?
2. Does RR predict ideational fluency, and how does the strength of that prediction change depending on participants' level of fluency?
3. Does RR predict originality, and how does the strength of that prediction change depending on participants' level of originality?
4. Does fluency predict RR, and how does the strength of that prediction change depending on participants' level of RR?
5. Does originality predict RR, and how does the strength of that prediction change depending on participants' level of RR?

6. Given the empirical findings related to questions 1–5, is there evidence to support the TH, or is another more nuanced explanation of the relations among these constructs required?

With these six empirical research questions formally posited, the methodology of the current study will now be overviewed.

2. Methodology and scoring

2.1. Participants

77 undergraduate students at a large public research university in the mid-Atlantic region of the United States, (58 female; 75.32%) participated in this study. Participants were recruited for this study via the introductory psychology participant pool, in which students taking introductory psychology classes are required to participate. In exchange for their participation, students received research-credit in their introductory psychology class. Participants ranged in age from 18 to 25, with a mean age of 19.83 ($SD = 1.44$). The sample was relatively diverse, although with a White majority (n = 50; 64.9%). 7.79% of students reported their ethnicity as African American (n = 6); 15.58% of students reported their ethnicity as Hispanic/Latino (n = 12); and 16.88% reported their ethnicity as Asian (n = 13). These ethnicity categories were not considered to be mutually exclusive, and participants were directed to select all that applied to them. The vast majority (89.61%) of the sample reported their first language as English (n = 69), and participants reported a mean grade point average (GPA) of 3.36 ($SD = 0.36$), with self-reported GPAs ranging from 2.50 to 4.00.

2.2. Measures

This study utilized two measures, the Alternate Uses Task (AUT) and the Test of Relational Reasoning (TORR). Each of these measures are now further explained.

2.2.1. Alternate uses task

The AUT is a commonly used psychometric task in which participants are asked to produce as many novel uses for a given object as possible. The AUT has been used for assessing participants' divergent thinking and creative ability for many years (Guilford, 1967; Hudson, 1968; Torrance, 1972), and remains one of the most-often utilized task within the creativity research literature (Puryear, Kettler, & Rinn, 2017). On the AUT, ten different object names were presented to participants in a random order. The ten objects that were used on the AUT were chosen based on the results of an empirical norming study, which identified them as objects that U.S. undergraduate students were generally very familiar with (Van Overshelde, Rawson, & Dunlosky, 2004). The object names that were presented to participants were as follows: book, fork, table, hammer, pants, trumpet, truck, carrot, shovel, and sandals.

One advantage of the AUT is that participant responses on these ten items provide information about both participants' ideational fluency and their originality. In previous measurement work (e.g., Dumas & Dunbar, 2014), these two dimensions of divergent thinking have been empirically shown to be distinct but correlated constructs. This study follows these recent methodological recommendations by quantifying participants' fluency and originality separately, and not combining them into a single score. To accomplish this, responses on each of the ten AUT objects were scored in two different ways.

2.2.1.1. Fluency scoring. In order to assess students' ideational fluency, the number of uses generated by each participant for each object was tallied. Then, the number of uses was summed across all ten objects, producing a total-uses variable for every participant. Counts such as these are the principle way in which fluency has been operationalized in the extant research literature (Plucker & Makel, 2010). In this study, participants' fluency scores exhibited a very high level of observed-variable scale reliability ($\alpha = 0.95$). Although such fluency counts have repeatedly been shown to be useful (e.g., Jauk et al., 2013; Karwowski et al., 2016), they do not provide information about the relative originality of participant responses. For that, another scoring mechanism is needed.

2.2.1.2. Originality scoring. In this study, originality was assessed using Latent Semantic Analysis (LSA) using a very large corpus (i.e., 37,651 included documents and more than 11 million included words) built to approximate the expected reading experience of the average college student (*general-reading-up-to-the-first-year-in-college*: Kintsch, 2000; Landauer et al., 1998). The semantic similarity of AUT object-prompts (e.g., Book) and each individual participant-generated use (e.g., read; throw like a Frisbee) was tested. As previously mentioned, LSA produces similarity statistics such that a relation of 1 indicates that two words or phrases are extremely similar and -1 indicates that they are totally dissimilar. In order to simplify the later analysis and make the indication of originality more straightforward, these similarity statistics were subtracted from 1, which yielded indicators of originality that ranged from 0 (not original) to 2 (very original). These originality scores were then averaged within each AUT object prompt, producing ten separate originality-scored items (i.e., each of the ten objects on the AUT). These ten scores were then summed to produce a total originality indicator across the AUT.

It should be noted that, within the creativity literature, many different originality scoring procedures exist (c.f., Plucker & Makel, 2010), but this study follows recently posited measurement recommendations based on construct reliability comparisons (Dumas & Dunbar, 2014). Moreover, LSA scores such as these have been shown to be a more reliable means of scoring the originality of student responses to the AUT than are panels of human raters (Forster & Dunbar, 2009), and in this study, originality across the ten AUT objects exhibited satisfactory scale reliability ($\alpha = 0.81$), which is actually relatively high for the creativity and originality literature (Dumas & Dunbar, 2014; Silvia, 2011). Another critical aspect of LSA originality scoring that must be noted here is that LSA similarity statistics, and therefore LSA originality scores, will necessarily differ depending on the corpus of text on which the LSA analysis is based. Therefore, the representational nature of the corpus, and the suitability of the corpus for use with a particular sample of individuals is of utmost importance. In this study, the corpus utilized (*general-reading-up-to-the-first-year-in-college;* Kintsch, 2000; Landauer et al., 1998) was constructed to be generalizable to the U.S. undergraduate population from which this study's sample was drawn, and is the most commonly used corpus within the divergent thinking literature (Dumas & Dunbar, 2014, 2016; Hass, 2017).

2.2.2. TORR

The TORR is a visuo-spatial RR measure with 32 items on four scales, each with 8 items, and each designed to tap one of the four forms of the construct (Alexander et al., 2015). In addition to these 32 scored items, each scale on the TORR features two sample items, designed to familiarize participants with the format of the scale, and on which automated feedback is given. Each TORR item is selected-response in format, has four possible answer choices, and is scored dichotomously (i.e. 0 = incorrect, 1 = correct). The four scales of the TORR are presented to participants in a random order, although the items within those scales have a fixed order based on the construction of their stimuli.

Although the TORR contains four distinct scales, a recent multidimensional item-response theory (MIRT) calibration and norming study (Dumas & Alexander, 2016) found that item-responses were best modeled using a bifactor model that included a general RR factor loading on all 32 items, and scale-specific residualized factors that explained scale-specific dependence among items. This same norming study also forwarded a specific conversion procedure for raw TORR responses, in which raw TORR scores are converted to normative scores—termed the *relational reasoning quotient* (RRQ)—that are based on the empirical Bayes scoring used in the MIRT context. Therefore, because RRQ scores are based on the scoring of the MIRT bifactor model, they are intended to indicate participant general RR ability without the problematic scale-specific dependency or variance. Following cognitive measurement tradition, the RRQ was standardized to have a mean of 100 and standard deviation of 15 within the norming undergraduate population, although sample-specific means in any given study will necessarily deviate from these values slightly. Further, if scale specific (e.g., analogy) scores are of interest, a similar normative scoring conversion is possible for scale-specific data, yielding scores with a normed mean of 25 and standard deviation of 5. Although it should be noted that, because of the nature of the MIRT models used to calibrate and norm the TORR, each normative scale score will not sum together to perfectly equal the RRQ.

In this study, RRQ scores are the main focus, but raw-to-normative score conversions were utilized for each individual scale of the TORR as well, for descriptive purposes. Because these conversions were used, the reported score reliability of the MIRT models used to create the normative scores by the test designers is relevant here. The RRQ converted scores have a reported marginal reliability (defined in the empirical Bayes MIRT scoring literature as 1 minus the averaged error variance divided by prior variance; Yang, Hansen, & Cai, 2012) of 0.81, which is relatively high in the visuo-spatial reasoning literature (Vock & Holling, 2008). The marginal reliabilities of normatively converted scale-specific scores are: analogy, 0.68; anomaly, 0.66; antinomy, 0.68; and antithesis, 0.70. It should be noted that, although these marginal reliabilities are necessarily lower than that of the RRQ because of the reduction in item number, MIRT analysis of the TORR (Dumas & Alexander, 2016) revealed that every TORR item was highly informative of its intended underlying construct at the latent variable level.

2.3. Procedures

Both the AUT and TORR were administered to participants online, via Qualtrics survey software. In order to access the measures, participants received a link from the introductory psychology participant-pool website. Once on the study website, participants completed an informed consent form, responded to both measures, and then provided demographic information. Participants were directed to complete the measures with minimal distractions, including turning off the TV, music, and closing other websites that may be open on their computer. Because the AUT in particular requires significant typing, participants were not permitted to participate via a smartphone or tablet, and were required to use a computer with a traditional keyboard.

In order to reduce the confounding effects of participant cognitive fatigue or stimuli ordering, the AUT and TORR were presented to participants in a randomized order. As previously mentioned, all items within the AUT, as well as the four scales of the TORR, were also presented in a random order. In this way, each participants' study experience was channeled through three different computerized randomizers. By calculating the permutations among these random elements, it can be seen that, cumulatively, these randomizers created approximately 174.2 million different possible distinct orders of the measure stimuli. Such a high number of permutations makes it very unlikely that two participants saw precisely the same stimuli order, greatly reducing the confounding effects of participant fatigue on any given measurement. It should also be noted that, in order to prevent the inflation of the relation between RR and DT because of participant speed-of-processing, one measure (i.e., AUT) was time-limited, while the other (i.e., TORR) was not. Specifically, participants were given two minutes to respond to each item on the AUT (for a total of 20 min on that measure), while no time-limit was imposed on the TORR.

Participation in this study occurred outside of the laboratory, from any computer connected to the Internet. For this study, Internet-based participation was considered beneficial, because it allowed for a beneficial level of participant flexibility and privacy. Moreover, previous research with the TORR (i.e., Alexander et al., 2015) showed that undergraduate scores on the online and paper version of the test were very similar, and that online participation may potentially improve testing fairness by limiting the anxiety that some participants feel when being assessed within a psychological laboratory. Moreover, creativity researchers have traditionally emphasized the need for participants to receive as much privacy as possible when responding to the AUT (e.g., Hudson, 1968), in order for them to feel comfortable providing responses that may otherwise be deemed inappropriate. Also, of course, Internet-based data collection has rapidly become the norm in many branches of educational and psychological research (e.g., Greene, 2009; Muenks, Miele, Ramani, Stapleton, & Rowe, 2015). With these procedures in place, participant data were collected and analyzed.

3. Results and implications

3.1. Score description

Means and standard deviations for each of the study variables are included in Table 1. As can be seen, participants produced approximately 7 possible uses per object on the AUT, and on average across each object, those uses were scored as 0.747 on the originality scale. Such a pattern is in line with previous LSA work (Dumas & Dunbar, 2014), and such originality scores make logical sense based on the LSA paradigm in which originality scores over 1 (i.e., latent semantic similarity statistics less than 0) are rare. Further, RRQ scores in this sample were very close to the normative mean of 100, as was each

Table 1
Descriptive statistics and bivariate correlations.

Variable	Mean	SD	Bivariate correlations						
			Originality	Fluency	Relational reasoning	Analogy	Anomaly	Antinomy	Antithesis
Originality	7.47	0.42	1.00						
Fluency	70.05	26.59	0.49**	1.00					
Relational Reasoning	98.41	10.86	0.32**	0.20	1.00				
Analogy	25.09	3.77	0.31**	0.14	0.78**	1.00			
Anomaly	24.30	3.34	0.23*	0.17	0.69**	0.44**	1.00		
Antinomy	25.32	4.73	0.27*	0.20	0.58**	0.23*	0.19	1.00	
Antithesis	24.11	3.83	0.03	0.02	0.67**	0.38**	0.32*	0.17	1.00

Note: Each scale of the TORR is normatively scored based on the norming population and item-response theory scoring procedure, and as such the means of each scale do not directly sum to the total score; When producing bivariate correlations, all variables were placed on a standard normal (M = 0, SD = 1) scale, therefore, these correlations are also equal to OLS regression coefficients with one predictor and one outcome.

** $p < .01$.

* $p < .05$.

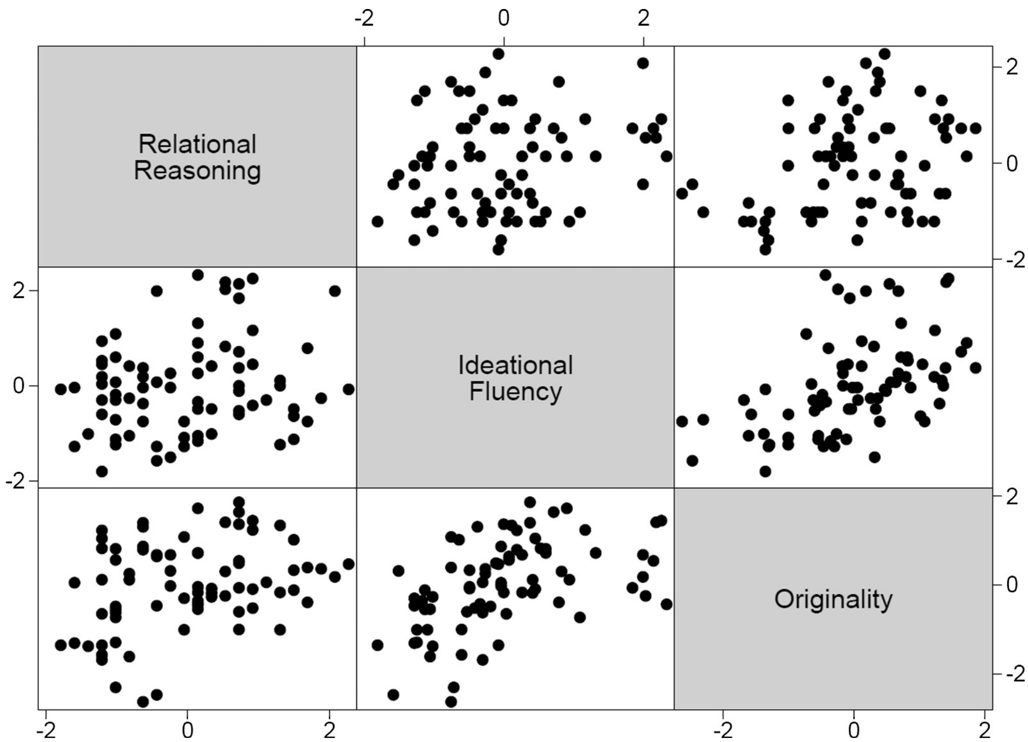

Fig. 1. Scatterplot matrix of bivariate relations among relational reasoning, ideational fluency, and originality. Note: All variables are standardized in this analysis.

individual scale normative score very close to its normative mean of 25. This pattern implies that the current sample is similar in terms of its RR ability to the sample used to norm the TORR. This observation is relevant to this investigation because some previous studies of the threshold hypothesis, utilizing different cognitive measures, have found sample means to be substantially higher than would be expected given test norms, possibly complicating the interpretation of study findings (e.g., Jauk et al., 2013).

3.2. Bivariate correlations

It should be noted here that, before calculating the bivariate correlations, the scores on the TORR and the AUT were placed on a standard normal (M = 0, SD = 1) metric. Although this choice does not affect the actual bivariate correlations among the scores, it allows a critical point of comparison with later analyses in this study. This is because, a bivariate correlation between two standardized variables is identical to the OLS regression coefficient when one standardized variable solely predicts the other. Also, because these variables are standardized, the regression coefficients are identical regardless of which variable is the predictor and which is the outcome. However, even with standardized variables, as previously discussed, quantile regression analysis does not exhibit such reciprocity, and ways in which the quantile regression results deviate from these bivariate correlations and from each other form the basis of critical inferences in this study. With that methodological choice in mind, the bivariate correlations among each of the study variables were compared (see Table 1).

Unsurprisingly based on past work (e.g., Dumas & Alexander, 2016), the strongest correlations depicted in Table 1 are among the TORR total score and the various scale-scores that encompass the TORR. In general, these scales tended to correlate together moderately strongly. Also relatively unsurprising was the moderately strong ($r = 0.49$) correlation between the two dimensions of divergent thinking, originality and fluency, which implies that participants who produced many ideas on the AUT on average were also likely to produce ideas that were more original on average (and vice versa).

More relevant to the research questions of this study is the observation that LSA-based originality scores correlated moderately strongly ($r = 0.32$) with the RRQ, leading to the conclusion that, in terms of a non-directional bivariate relationship, originality and RR are significantly positively related. This implies that those students who are more adept relational thinkers on average are also more capable of producing relatively original ideas on average (and vice versa). Originality was similarly significantly positively related to each of the TORR scales, except for antithesis. Such a finding may imply that antithetical thinking skill is less required for the production of original ideas than are the other forms of RR.

Although fluency scores were positively related to the RRQ and to each of the scales of the TORR, none of these bivariate correlations reached significance. This finding implies that, at least on average, those students who scored highly on the TORR did not necessarily produce more ideas on the AUT. This finding is actually in line with previous research in the creativity literature (e.g., Hass, 2017), in which the precise cognitive mechanisms by which participants produce more or fewer ideas during divergent thinking is not entirely understood. Perhaps ironically, fluency, despite being relatively straightforward to measure, is difficult to explain in the cognitive literature, and may be better accounted for—at least at the mean level—by other individual differences such as personality, culture, or mood (e.g., Dumas & Dunbar, 2016; Lewis & Lovatt, 2013). Of course, just because the non-directional bivariate correlations among fluency and RR are not significant, that does not imply that the directional regression coefficients between these variables are not significant for particular subsets (i.e., quantiles) of students.

In order to visually represent these bivariate relations, the three main variables in this study (i.e., relational reasoning, fluency, and originality) are featured in a scatterplot matrix in Fig. 1. As can be seen in this matrix, the linear relations among fluency and originality, and relational reasoning and originality, are stronger than that between RR and fluency, although all relations are positive. It should be noted that individual scale scores from the TORR are not included in this plot, or in the rest of the analysis, for a number of reasons. First, with only 8 items, the scale scores do not feature enough levels (i.e., they do not closely enough approximate continuousness) to be featured as outcomes in sequential quantile regression analysis (Koenker, 2005). Second, previous measurement work (e.g., Dumas & Alexander, 2016)

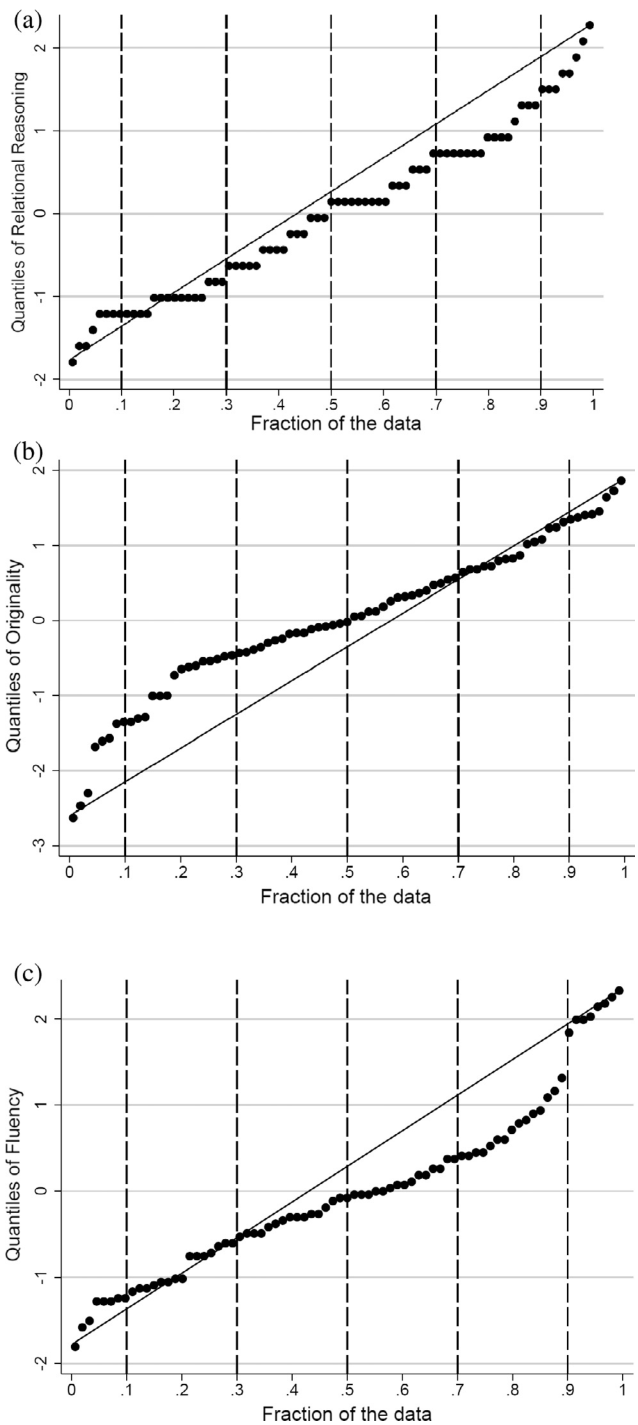

Fig. 2. Quantile plots for: (a) relational reasoning, (b) originality, and (c) ideational fluency.

has shown, and the bivariate correlations from this study concur, that the general score is significantly positively, and even strongly, related to each of its constituent scales.

3.3. Quantile description

Before beginning the quantile regression analysis, it is necessary to inspect the quantile distributions of each of the main variables in this study. Quantile plots for RR, originality, and fluency are available in Fig. 2. These plots are constructed by placing the standardized variable (based on the mean and standard deviation) on the vertical Y axis, and the quantiles, or individual fractions of the data, on the horizontal X axis. In this way, the distribution of the standardized variable across its quantiles can be visually observed. In order to better depict the rate at which the standardized variable changes across the quantiles, a reference line was added to these quantile plots by simply connecting the lowest data point to the highest data point. Vertical (slashed) reference lines are also added at the particular quantiles that will be featured in later analysis (0.1, 0.3, 0.5, 0.7, and 0.9).

Fig. 2a depicts the quantiles of RR. It should be noted that because this variable only has 32 possible levels, the quantile plot has a slightly step-like appearance. However, 32 levels is generally considered enough to model quantile-level relations among variables within the methodological literature (Koenker, 2005). This quantile plot shows that only a small number of students received the very lowest scores on the TORR, but that, beginning at the first decile (0.1), many more participant scores are present. Between the first and third deciles, only 3 different TORR scores are present, but as the plot moves closer to the median (5th quantile; 0.5), the scores appear to vary more. At the upper end of the RR distribution, a small number of participants received top scores, giving the plot a steeper rise beginning at the ninth decile (0.9).

In Fig. 2b, which holds the quantile plot for originality, it can be seen that a relatively small number of participants generated the most unoriginal uses, leading to a quantile plot that appears somewhat disjointed at the lowest quantiles. But beginning around the second decile (0.2), the improvement in originality across quantiles is more uniform, with only a slight uptick in participant originality scores after the ninth decile. It should be noted that because the LSA-based originality measure used in this study yields an entirely continuous score, the quantile plot for originality does not appear step-like, as does the quantile plot associated with RR.

Similarly to originality, fluency scores in this study are also a continuous measure, yielding a quantile plot (Fig. 2c) that does not appear step-like. However, in contrast to originality, the fluency quantile plot reveals a distribution of scores that improve at a near-uniform rate across the lower quantiles, with a rapid jump in scores only occurring at the 9th decile. Thus, each of the three main variables in this study have a meaningfully different distribution of scores across their quantiles. This observation requires that the predictive relation among the variables also differs across the quantiles, such that one variable may predict the other very strongly at one quantile, but not at another. The following quantile regression analysis seeks to uncover just such possible patterns.

3.4. Quantile regression results

This quantile regression analysis uses a predictor variable (x) to predict an outcome variable (y), at different quantiles *of the outcome variable* (Yu, Lu, & Stander, 2003). As opposed to ordinary-least-squares (OLS) regression, which predicts the mean of the outcome and is estimated by minimizing the magnitude of squared deviation terms, quantile regression predicts the median (or any other specified quantile) of the outcome and is estimated by minimizing the absolute value of the deviation terms (least-absolute-value [LAV] estimator; Gentle, 1977). In this way, although the estimating algorithm differs from traditional OLS regression analysis, interpretation of the regression coefficients does not substantively change in quantile regression, except that the coefficients pertain to particular specified quantiles of the outcome variable. Based on relevant past work, and in an effort to provide meaningful information across the distribution of each outcome variable, the following quantiles were chosen for this analysis: 0.1, 0.3, 0.5, 0.7, and 0.9. Therefore, the strength of the predictive relation between each predictor and outcome was tested at the first, third, fifth (median), seventh, and ninth deciles of the outcome variable. This analysis was performed in Stata version 13.1 (StataCorp, 2013) using the `sqreg` function. Standard errors were calculated through bootstrapping with 20 replications. It should be noted in this analysis that the statistical power to detect a true effect in quantile regression varies depending on the actual quantiles being tested in the model (Yu et al., 2003). In this way, the power to detect an effect at the lowest decile of a given distribution will likely be lower than the power to detect an effect at the median, because fewer individuals in a sample typically fall at the outer quantiles, and more fall in the middle of the distribution. Given this state-of-affairs, the significance of quantile regression coefficients in this analysis is discussed as directly pertaining to the quantile plots and description presented above. In addition, the bootstrap procedure conducted here for the estimation of standard errors and confidence intervals follows current methodological recommendations for validly maximizing the power of quantile regression procedures in relatively small samples such as the one analyzed here (Tarr, 2012).

Quantile regression results are here organized into two sections: (a) analysis pertaining to the relation between RR and originality, and (b) analysis pertaining to the relation between RR and fluency. As far as I am aware, the TH has not before been explicitly tested in both of these predictive directions. However, recent empirical work has found statistically and practically significant effects both when creative attributes predict cognitive and academic outcomes (e.g., Gajda et al., 2016) as well as when cognitive attributes predict creative outcomes (e.g., Karwowski, Kaufman, Lebuza, Szumski, & Firkowska-Mankiewicz, 2017). In addition, from a neuropsychological perspective, there is extant evidence that creative abilities such as DT support students' ability to solve complex cognitive problems, including those that require relational reasoning (Green et al., 2017), and therefore cognitive abilities do not necessarily precede (i.e., as an independent variable) creative attributes. Methodologically, quantile regression allows for the TH to be examined in both of these predictive directions. This is because the individual distributions of each variable (i.e., independent and dependent variables) at each quantile being modeled is taken into account. This means that, if the observed distributions of the independent and dependent variables are not exactly the same (a situation that would be highly unlikely) then even when the variables are standardized and there is only one predictor, the coefficients estimated in a quantile regression model differ depending on which variable is the predictor and which is the outcome. Therefore, in order to provide the richest possible explanation of the relations among these variables, each variable was tested as both the predictor and the outcome in this analysis.

3.4.1. Originality

Table 2 contains quantile regression coefficients by which RR predicts originality. As can be seen, RR is a significant, and relatively strong ($\beta = 0.54$, $p < .01$) predictor of originality at the first decile, and is also a significant predictor at the third decile and median. However, the strength of this relation wanes as the quantiles increase, eventually becoming non-significant and essentially zero at the seventh decile, before slightly regaining strength (but not significance) at the ninth decile. Such a pattern of coefficients implies that RR is critically related to original thinking ability, but only for those individuals who are at the median or below on originality. Indeed, this pattern may support the threshold hypothesis, suggesting that RR ability is associated with original thinking, but only to the median level.

However, this pattern is complicated when the predictive direction is reversed such that originality is the predictor and RR is the outcome (see Table 3). In this case, originality is a significant predictor of RR in the third, fifth, and seventh deciles, but not the first or ninth. However, the magnitude of the regression coefficient for the ninth decile is actually the strongest across all the quantiles ($\beta = 0.38$, $p = .13$), but the small number of participants who fall in that quantile meant that the standard error was large and therefore the coefficient was not significant. Additionally, the coefficient at the first decile, although weakest across the quantiles, was only barely non-significant ($\beta = 0.17$, $p = .05$). Therefore, the predictive relation from originality to RR seems relatively strong and positive across the distribution of RR. Such a pattern of coefficients implies that those participants who were highly skilled at reasoning on the TORR were also likely to produce very original ideas on the AUT.

Table 3
Quantile regression results wherein originality predicts relational reasoning.

Quantile of relational reasoning	Intercept	Coefficient	S.E.	t	p
0.10	−1.16	0.17	0.09	1.9	.05
0.30	−0.67	0.34**	0.12	2.84	< .01
0.50	0.05	0.35**	0.08	4.30	< .01
0.70	0.42	0.34**	0.09	3.66	< .01
0.90	1.36	0.38	0.25	1.51	.13

Note: Variables are standardized (M = 0, SD = 1) for this analysis.
* $p < .05$.
** $p < .01$.

To visualize the pattern of differences among these predictive relations, the regression coefficients from these two sequential quantile regression models are plotted in Fig. 3. In this figure, the dotted horizontal reference line depicts the bivariate correlation between originality and RR. As can be seen, in the first and third deciles, RR better predicts originality than vice versa, but beginning at the median, that relation is reversed such that originality better predicts RR. This overarching pattern illustrates that, for those participants who were not very original, RR was associated with their originality, but for those who were highly original, RR was not strongly associated with their originality. Taken together, this implies that RR is necessary but not sufficient for original thinking, a finding that is in keeping with the threshold hypothesis generally.

3.4.2. Fluency

Table 4 presents quantile regression coefficients for RR predicting ideational fluency. The predictive strength of this relation is near-zero until the seventh decile of fluency, when it increases but remains non-significant ($\beta = 0.24$, $p = .28$). Only at the ninth decile of fluency does RR become a significant predictor ($\beta = 0.62$, $p < .01$), implying that RR's association with fluency is confined to those participants who produce very many ideas on the AUT.

In contrast, nearly the opposite pattern is observed when fluency predicts RR (Table 5). Specifically, fluency only significantly predicts RR at the first and second decile of RR ability, with weakening non-significant influence in the fifth and seventh deciles. At the ninth decile fluency negatively (although non-significantly) predicts RR. This pattern of coefficients suggests that only among the first and second deciles of fluency were those participants who were more fluent also more capable at RR.

As with the previous analysis, both of these sequences of regression coefficients are visually plotted in Fig. 4, with the horizontal dotted reference line representing the bivariate correlation between fluency and RR. As the plot shows, from the first decile through the median, fluency is a better predictor of RR than vice versa. However, beginning at the seventh decile and greatly accentuated at the ninth, RR is a stronger predictor of fluency than vice versa. Overall, this pattern

Table 2
Quantile regression results wherein relational reasoning predicts originality.

Quantile of originality	Intercept	Coefficient	S.E.	t	p
0.10	−1.02	0.54**	0.09	5.61	< .01
0.30	−0.48	0.42**	0.09	4.41	< .01
0.50	−0.18	0.29*	0.12	2.36	.021
0.70	0.61	−0.06	0.16	−0.38	.71
0.90	1.27	0.18	0.15	1.19	.24

Note: Variables are standardized (M = 0, SD = 1) for this analysis.
* $p < .05$.
** $p < .01$.

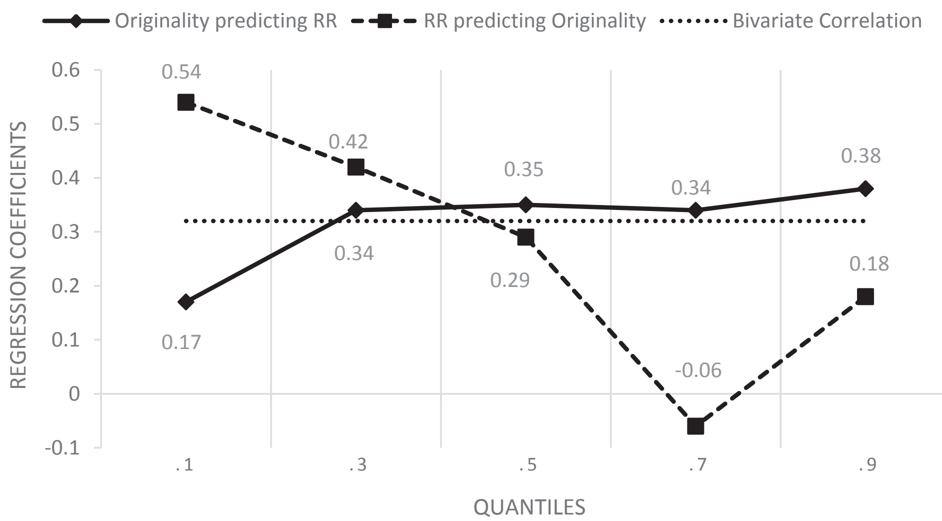

Fig. 3. Plot of quantile regression coefficients pertaining to relational reasoning and originality. Note: All variables are standardized in this analysis; dotted reference line represents bivariate correlation (r = 0.32).

Table 4
Quantile regression results wherein relational reasoning predicts ideational fluency.

Quantile of fluency	Intercept	Coefficient	S.E.	*t*	*p*
0.10	−1.17	0.06	0.16	0.41	.68
0.30	−0.61	0.11	0.13	0.81	.42
0.50	−0.08	0.07	0.15	0.01	.99
0.70	0.36	0.24	0.23	1.08	.28
0.90	1.68	0.62**	0.15	4.07	< .01

Note: Variables are standardized (M = 0, SD = 1) for this analysis.
*p < .05.
** p < .01.

Table 5
Quantile regression results wherein ideational fluency predicts relational reasoning.

Quantile of relational reasoning	Intercept	Coefficient	S.E.	*t*	*p*
0.10	−1.21	0.18*	0.07	2.30	.02
0.30	−0.68	0.29*	0.13	2.19	.03
0.50	−0.03	0.24	0.14	1.70	.09
0.70	0.44	0.21	0.13	1.53	.13
0.90	1.32	−0.15	0.28	−0.55	.54

Note: Variables are standardized (M = 0, SD = 1) for this analysis.
**p < .01.
* p < .05.

suggests that those participants that were very high (i.e. ninth decile) on fluency were also likely to be capable relational reasoners, but those participants who were very high on RR were actually unlikely to be highly fluent. This pattern differs substantially from that found between RR and originality, and does not appear to support the threshold hypothesis. Therefore, taken together, the cumulative findings from this investigation may require a modification of existing theory to explain.

4. Discussion and conclusion

RR and DT are two critical antecedents of creative problem solving, but their interrelations are not currently well-understood. This study has been the first to use quantile regression to uncover the varying relations among dimensions of DT (i.e., originality and fluency) and RR, across the distributions of each of those variables. Although a number of previous investigations have empirically tested the TH (e.g., Jauk et al., 2013; Preckel et al., 2006) this work provides finer-grain information about this hypothesis than has previously been available. In this way, this study holds a number a principal findings that are relevant to the TH, as well as the field's broader educational and psychological understanding of RR and DT. Here, those principal findings are organized under two broad headings: (a) the relation between RR and originality and (b) the relation between RR and fluency. After these findings are presented, their educational implications are discussed.

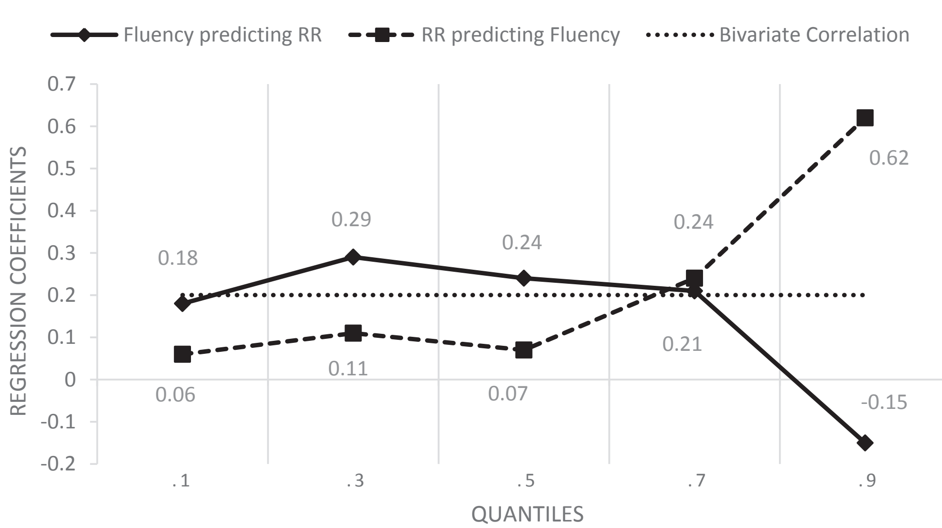

Fig. 4. Plot of quantile regression coefficients pertaining to relational reasoning and ideational fluency. Note: All variables are standardized in this analysis; dotted reference line represents bivariate correlation (r = 0.20).

4.1. Relational reasoning and originality: support for the threshold hypothesis

At the mean level, originality was significantly linearly related (i.e., bivariate correlations) to RR and each of its forms except antithesis. This finding suggests that those students who are more capable at RR on average are also more likely to produce more original responses on average to the AUT. However, as hypothesized, this relation was complicated when the predictive relation between the variables were estimated at individual quantiles of each construct. Specifically, RR was only significantly predictive of originality at and below the median of originality, although originality was more uniformly predictive of RR. Based on this observation, it appears that RR is a necessary-but-not sufficient condition for originality, in that, for those students at or below the median of originality, those with greater RR ability are likely to also be more original. However, for those students above the median of originality, those with greater RR ability are no more likely to be highly original. In this way, following Guilford's (1967) threshold hypothesis, RR is significantly supportive of originality, but only up to a point.

However, the findings of this study do differ from the traditional TH in a number of ways. For one, the threshold uncovered here is actually on the distribution of the creative ability (i.e. originality) not the cognitive (i.e. RR). So, whereas Guilford (1967) and others (e.g., Cho et al., 2010; Kim, 2005) have argued for a threshold on the distribution of cognition, this study suggests that RR and originality are relatively similarly related across the distribution of RR, but their relations differ substantially across the distribution of originality. Such a finding lends credence to the methodological and substantive argument forwarded in this article, that the TH should be tested in both hypothesized directions, instead of just one. Secondly, the threshold here identified was located essentially at the median of originality, which is a very different location from the traditional IQ = 120 location typically utilized in the existing literature (e.g., Sligh et al., 2005). Indeed, this finding follows others (Karwowski et al., 2016) who have shown that the IQ = 120 threshold is too simplistic to describe reality. In sum, the findings of this study associated with RR and originality do support the TH, albeit with some caveats including the variable on which the threshold is located (i.e., originality instead of RR) and the location of that threshold (i.e., the median of originality instead of RRQ = 120).

4.2. Relational reasoning and fluency: evidence against the threshold hypothesis

In contrast to the findings of this investigation related to RR and originality, which generally supported the substantive core of the TH, the findings related to RR and fluency require greater theoretical modification to thoroughly explain. First, at the mean level, fluency did not significantly correlate with RR or any of its forms. This finding is noticeably different from other published studies that have found stronger relations between other RR measures (e.g., RPM) and fluency (e.g., Shi et al., 2017). However, significant positive relations between RR and fluency did emerge in this study during quantile regression analysis. However, these quantile-specific relations between RR and fluency did not coincide to those quantile-specific relations hypothesized by the TH. Specifically, while RR was only predictive of originality in the lower-half of the distribution of originality, RR was only predictive of fluency at the very top (i.e., 9th decile) of the fluency distribution. Moreover, the coefficient at that point was the strongest estimated in this study (i.e. $\beta = 0.62$) implying that RR critically supports ideational fluency, but only for those students who produce the very most ideas on the AUT. Moreover, fluency was only significantly predictive of RR at or below the third decile of the RR distribution, such that only for those participants who performed relatively poorly on the TORR did a higher level of fluency lead to better RR. Given these results, a tentative threshold in the relation between fluency and RR may be discerned at the third decile of RR ability, although the general findings (including those related to originality) require a mechanistic hypothesis beyond the TH to explain.

In my view, these findings taken together imply that capability in RR does allow students to increase their ideational fluency, although only at very high levels of fluency, perhaps because a high level of disinhibition in AUT responses is required for the benefits of greater RR ability to be observed on ideation (Radel, Davranche, Fournier, & Dietrich, 2015). Such a relation is analogous to those found within economics research using quantile regression, in which a given individual variable improves financial outcomes more for individuals who are already doing well. A similar pattern was uncovered by Dumas and Schmidt (2015), in which RR ability predicted engineering students' ability to benefit from instruction, such that those students who exhibited high RR ability tended to develop other abilities (including originality) as well.

4.3. Implications for educational intervention

Given the critical societal need to prepare students to creatively solve the problems of the future, intervention on identified antecedents of CPS may be interesting to a number of educational psychologists. Importantly, some of the value of such interventions is derived from the expected outcome that intervention on one mental ability (e.g., RR) may cause shifts in other abilities (e.g., DT). Nonetheless, previous to this study, expected changes on one of these variables (i.e. RR or DT) based on growth in another could not be predicted in a meaningful way. However, based on the findings of this study, some expectations can be forwarded. For example, intervening on students' RR ability may lead to associated improvements on their originality, but only for those students who are at or below the median of originality (at least among undergraduate students). Therefore, such intervention may be most effective within groups of students who have been previously identified as moderate to low original thinkers, or who are expected to be so based on some other individual difference (e.g., age). In contrast however, intervention on RR is predicted based on the findings of this study to be most effective in improving ideational fluency for those students who are already highly fluent (i.e., 9th decile of fluency). In this way, RR intervention would be expected to be most effective at improving DT in students who are simultaneously exhibit low-originality and high fluency. In addition, improvements to students' originality is predicted, based on these findings, to improve RR across most of its distribution. Notably, originality has been shown in the extant literature to be malleable within even very short motivational or mood-based interventions (e.g., Dumas & Dunbar, 2016; Oppezzo & Schwartz, 2014), and the current findings imply that these simple methods may also be effective at supporting RR, a hypothesis that generally remains to be tested.

Moreover, any improvements on ideational fluency gained from an intervention would be hypothesized, based on the findings of this study, to lead to gains in RR ability only for those students who are at or below the third decile of RR ability. Taken together, DT interventions may be least appropriate for students who are already adept at RR, but effective for those who are in the middle or lower portions of the distribution of that variable.

One major caveat to any discussion of the malleability and instruction of generalized mental abilities such as RR or DT is that, from a cognitive neuroscience perspective, improvement may only be possible within a particular domain of learning and not in a truly general way (Sala & Gobet, 2017). Indeed, one major finding of the educational and psychological literatures has been that the transfer of gains on one measure or construct to other tasks, constructs, or domains, is often very limited (Sprenger et al., 2013), although at least one major meta-analysis did find positive evidence for the transfer of learned problem-solving strategies (Klauer & Phye, 2008). Given this evidence, the most fruitful way to proceed may be to nest instructional interventions focusing on either RR or DT within particular domains of learning, and compliment the RR or DT training with instruction of necessary domain knowledge to support CPS.

Through the use of quantile regression methods, this study has allowed for a much finer-grained look at the relations among DT and RR than have previous investigations. Such a nuanced psychological understanding of these constructs and the way they interact is critical for the future planning of educational interventions, or for the eventual inclusion of creative or relational curriculum in schools. Indeed, it is hoped that as the field's understanding of DT and RR continues to grow, that knowledge will be critically useful in accomplishing a fundamental goal of 21st century education: meaningful support for all students' creative problem solving.

Author note

The author declares that they have no conflicts of interest.

References

Acar, S., & Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal, 26*(2), 229–238. http://dx.doi.org/10.1080/10400419.2014.901095.

Acar, S., & Runco, M. A. (2017). Latency predicts category switch in divergent thinking. *Psychology of Aesthetics, Creativity, and the Arts, 11*(1), 43–51. http://dx.doi.org/10.1037/aca0000091.

Alban-Metcalfe, R. J. (1978). Divergent thinking "threshold effect": IQ, age, or skill? *Journal of Experimental Education, 47*(1), 4–8. http://dx.doi.org/10.1080/00220973.1978.11011647.

Alexander, P. (2012). Reading into the future: Competence for the 21st century. *Educational Psychologist, 47*(4), 259–280. http://dx.doi.org/10.1080/00461520.2012.722511.

Alexander, P. A., Dumas, D., Grossnickle, E. M., List, A., & Firetto, C. M. (2015). Measuring relational reasoning. *Journal of Experimental Education, 84*(1), 119–151. http://dx.doi.org/10.1080/00220973.2014.963216.

Alfonso-Benlliure, V., & Santos, M. R. (2016). Creativity development trajectories in elementary education: Differences in divergent and evaluative skills. *Thinking Skills and Creativity, 19*, 160–174. http://dx.doi.org/10.1016/j.tsc.2015.11.003.

An, D., Song, Y., & Carr, M. (2016). A comparison of two models of creativity: Divergent thinking and creative expert performance. *Personality and Individual Differences, 90*, 78–84. http://dx.doi.org/10.1016/j.paid.2015.10.040.

Baldo, J. V., Bunge, S. A., Wilson, S. M., & Dronkers, N. F. (2010). Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain and Language, 113*(2), 59–64. http://dx.doi.org/10.1016/j.bandl.2010.01.004.

Beaty, R. E., Christensen, A. P., Benedek, M., Silvia, P. J., & Schacter, D. L. (2017). Creative constraints: Brain activity and network dynamics underlying semantic interference during idea production. *NeuroImage, 148*, 189–196. http://dx.doi.org/10.1016/j.neuroimage.2017.01.012.

Braasch, J. L. G., & Goldman, S. R. (2010). The role of prior knowledge in learning from analogies in science texts. *Discourse Processes, 47*(6), 447–479. http://dx.doi.org/10.1080/01638530903420960.

Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology, 94*(2), 327–343. http://dx.doi.org/10.1037/0022-0663.94.2.327.

Cho, S. H., Nijenhuis, J. T., van Vianen, A. E. M., Kim, H. B., & Lee, K. H. (2010). The relationship between diverse components of intelligence and creativity. *The Journal of Creative Behavior, 44*(2), 125–137. http://dx.doi.org/10.1002/j.2162-6057.2010.tb01329.x.

Choi, J. Y., Elicker, J., Christ, S. L., & Dobbs-Oates, J. (2016). Predicting growth trajectories in early academic learning: Evidence from growth curve modeling with Head Start children. *Early Childhood Research Quarterly, 36*, 244–258. http://dx.doi.org/10.1016/j.ecresq.2015.12.017.

Coll, R. K. (2008). Effective chemistry analogies. In A. G. Harrison, & R. K. Coll (Eds.). *Using analogies in middle and secondary science classrooms: The FAR guide – An interesting way to teach with analogies* (pp. 127–174). Thousand Oaks, CA, US: Corwin Press.

Danielson, R. W., & Sinatra, G. M. (2016). A relational reasoning approach to text-graphic processing. *Educational Psychology Review*. http://dx.doi.org/10.1007/s10648-016-9374-2.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science and Technology, 41*, 391–407.

Dumas, D., & Alexander, P. A. (2016). Calibration of the test of relational reasoning. *Psychological Assessment, 28*(10), 1303–1318. http://dx.doi.org/10.1037/pas0000267.

Dumas, D., Alexander, P. A., Baker, L. M., Jablansky, S., & Dunbar, K. N. (2014). Relational reasoning in medical education: Patterns in discourse and diagnosis. *Journal of Educational Psychology, 106*, 1021–1035. http://dx.doi.org/10.1037/a003677.

Dumas, D., Alexander, P. A., & Grossnickle, E. M. (2013). Relational reasoning and its manifestations in the educational context: A systematic review of the literature. *Educational Psychology Review, 25*(3), 391–427. http://dx.doi.org/10.1007/s10648-013-9224-4.

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

Dumas, D., & Dunbar, K. N. (2016). The creative stereotype effect. *PLoS ONE, 11*(2), e0142567. http://dx.doi.org/10.1371/journal.pone.0142567.

Dumas, D., & Schmidt, L. (2015). Relational reasoning as predictor for engineering ideation success using TRIZ. *Journal of Engineering Design, 26*(1–3), 74–88. http://dx.doi.org/10.1080/09544828.2015.1020287.

Dumas, D., Schmidt, L. C., & Alexander, P. A. (2016). Predicting creative problem solving in engineering design. *Thinking Skills and Creativity, 21*, 50–66. http://dx.doi.org/10.1016/j.tsc.2016.05.002.

Ehri, L. C., Satlow, E., & Gaskins, I. (2009). Grapho-phonemic enrichment strengthens keyword analogy instruction for struggling young readers. *Reading and Writing Quarterly: Overcoming Learning Difficulties, 25*(2–3), 162–191. http://dx.doi.org/10.1080/10573560802683549.

Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques.* Baltimore: University Park Press.

Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology, 45*(4), 554–558. http://dx.doi.org/10.1111/j.1469-8986.2008.00656.x.

Foltz, P. W., Streeter, L. A., Lochbaum, K. E., & Landauer, T. K. (2013). Implementation and applications of the intelligent essay assessor. In M. D. Shermis, & J. Burstein (Eds.). *Handbook of automated essay evaluation: Current applications and new directions* (pp. 68–88). New York, NY, US: Routledge/Taylor & Francis Group.

Forster, E. A., & Dunbar, K. N. (2009). Creativity evaluation through latent semantic analysis. *Paper presented at the 2009 annual meeting of the Cognitive Science Society, Amsterdam, Netherlands, July 29–August 1, 2009*.

Furnham, A. (2016). Individual differences in intelligence, personality and creativity. In J. C. Kaufman, & J. Baer (Eds.). *Creativity and reason in cognitive development* (pp. 327–353). New York, NY, US: Cambridge University Press.

Gajda, A., Karwowski, M., & Beghetto, R. A. (2016). Creativity and academic achievement: A meta-analysis. *Journal of Educational Psychology*. http://dx.doi.org/10.1037/edu0000133.

Gentle, J. E. (1977). Least absolute values estimation: an introduction. *Communications in Statistics – Simulation and Computation, 6*(4), 313–328. http://dx.doi.org/10.1080/03610917708812047.

Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2010). Connecting long distance: Semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cerebral Cortex, 20*(1), 70–76. http://dx.doi.org/10.1093/cercor/bhp081.

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*(2), 264–272. http://dx.doi.org/10.1037/a0025764.

Green, A. E., Spiegel, K. A., Giangrande, E. J., Weinberger, A. B., Gallagher, N. M., & Turkeltaub, P. E. (2017). Thinking cap plus thinking zap: tDCS of frontopolar cortex improves creative analogical reasoning and facilitates conscious augmentation of state creativity in verb generation. *Cerebral Cortex, 27*(4), 2628–2629.

Greene, J. A. (2009). Collegiate faculty expectations regarding students' epistemic and ontological cognition and the likelihood of academic success. *Contemporary Educational Psychology, 34*(3), 230–239. http://dx.doi.org/10.1016/j.cedpsych.2009.05.003.

Grossnickle, E. M., Dumas, D., Alexander, P. A., & Baggetta, P. (2016). Individual differences in the process of relational reasoning. *Learning and Instruction, 42*(Suppl. C), 141–159. http://dx.doi.org/10.1016/j.learninstruc.2016.01.013.

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

Hao, N., Liu, M., Ku, Y., Hu, Y., & Runco, M. A. (2015). Verbal divergent thinking facilitated by a pleasurable incubation interval. *Psychology of Aesthetics, Creativity, and the Arts, 9*(3), 286–295. http://dx.doi.org/10.1037/a0038851.

Hardy, J. H. I., & Gibson, C. (2017). Gender differences in the measurement of creative problem-solving. *The Journal of Creative Behavior, 51*(2), 153–162. http://dx.doi.org/10.1002/jocb.92.

Hargrove, R. A., & Nietfeld, J. L. (2015). The impact of metacognitive instruction on creative problem solving. *Journal of Experimental Education, 83*(3), 291–318. http://dx.doi.org/10.1080/00220973.2013.876604.

Hass, R. W. (2017). Semantic search during divergent thinking. *Cognition, 166*, 344–357. http://dx.doi.org/10.1016/j.cognition.2017.05.039.

Hennessey, B. A. (2010). The creativity—Motivation connection. In J. C. Kaufman, & R. J. Sternberg (Eds.). *The Cambridge handbook of creativity* (pp. 342–365). New York, NY, US: Cambridge University Press.

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak, & R. G. Morrison (Eds.). *The oxford handbook of thinking and reasoning* (pp. 234–259). New York: Oxford University Press.

Hudson, L. (1968). *Frames of mind: Ability, perception and self-perception in the arts and sciences.* Oxford, England: W.W. Norton.

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. http://dx.doi.org/10.1016/j.intell.2013.03.003.

Jones, L. L., & Estes, Z. (2015). Convergent and divergent thinking in verbal analogy. *Thinking & Reasoning, 21*(4), 473–500. http://dx.doi.org/10.1080/13546783.2015.1036120.

Karwowski, M., Dul, J., Gralewski, J., Jauk, E., Jankowska, D. M., Gajda, A., ... Benedek, M. (2016). Is creativity without intelligence possible? A necessary condition analysis. *Intelligence, 57*, 105–117. http://dx.doi.org/10.1016/j.intell.2016.04.006.

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

Karwowski, M., Kaufman, J., Lebuza, I., Szumski, G., & Firkowska-Mankiewicz, A. (2017). Intelligence in childhood and creative achievements in middle-age: The necessary condition approach. *Intelligence, 64*, 36.

Kassim, H., Nicholas, H., & Ng, W. (2014). Using a multimedia learning tool to improve creative performance. *Thinking Skills and Creativity, 13*, 9–19. http://dx.doi.org/10.1016/j.tsc.2014.02.004.

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). New York, NY, US: Cambridge University Press.

Kendeou, P., Butterfuss, R., Van Boekel, M., & O'Brien, E. J. (2016). Integrating relational reasoning and knowledge revision during reading. *Educational Psychology Review*. http://dx.doi.org/10.1007/s10648-016-9381-3.

Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education, 16*(2–3), 57–66.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review, 7*, 257–266.

Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of Educational Research, 78*(1), 85–123. http://dx.doi.org/10.3102/0034654307313402.

Koenker, R. (2005). *Quantile regression.* NY: Cambridge University Press.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25*, 259–284.

Language and Reading Research Consortium, & Logan, J. (2017). Pressure points in reading comprehension: A quantile multiple regression analysis. *Journal of Educational Psychology, 109*(4), 451–464. http://dx.doi.org/10.1037/edu0000150.

Lewis, C., & Lovatt, P. J. (2013). Breaking away from set patterns of thinking: Improvisation and divergent thinking. *Thinking Skills and Creativity, 9*, 46–58. http://dx.doi.org/10.1016/j.tsc.2013.03.001.

Mayer, R. E. (2016). The role of domain knowledge in creative problem solving. In J. C. Kaufman, & J. Baer (Eds.). *Creativity and reason in cognitive development* (pp. 147–163). New York, NY, US: Cambridge University Press.

McNeish, D., & Dumas, D. (2017). Nonlinear growth models as measurement models: A second-order growth curve model for measuring potential. *Multivariate Behavioral Research, 52*(1), 61–85. http://dx.doi.org/10.1080/00273171.2016.1253451.

Mohr, A. H., Sell, A., & Lindsay, T. (2016). Thinking inside the box: Visual design of the response box affects creative divergent thinking in an online survey. *Social Science Computer Review, 34*(3), 347–359. http://dx.doi.org/10.1177/0894439315588736.

Muenks, K., Miele, D. B., Ramani, G. B., Stapleton, L. M., & Rowe, M. L. (2015). Parental beliefs about the fixedness of ability. *Journal of Applied Developmental Psychology, 41*, 78–89. http://dx.doi.org/10.1016/j.appdev.2015.08.002.

Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine, 22*(19), 3055–3071.

Oppezzo, M., & Schwartz, D. L. (2014). Give your ideas some legs: The positive effect of walking on creative thinking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(4), 1142–1152. http://dx.doi.org/10.1037/a0036577.

Orzechowski, J., Kruchowska, E., Gruszka, A., & Szymura, B. (2017). Understanding factors behind the effectiveness of personal identification: Revolution—A new technique of creative problem solving. *Thinking Skills and Creativity, 23*, 140–149. http://dx.doi.org/10.1016/j.tsc.2016.12.004.

Parnes, S. J. (1961). Effects of extended effort in creative problem solving. *Journal of Educational Psychology, 52*(3), 117–122. http://dx.doi.org/10.1037/h0044650.

Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. C. Kaufman, & R. J. Sternberg (Eds.). *The Cambridge handbook of creativity* (pp. 48–73). New York, NY, US: Cambridge University Press.

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*(1), 159–170. http://dx.doi.org/10.1016/j.paid.2005.06.022.

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*(1), 59–68. http://dx.doi.org/10.1037/aca0000079.

Radel, R., Davranche, K., Fournier, M., & Dietrich, A. (2015). The role of (dis)inhibition in creativity: Decreased inhibition improves idea generation. *Cognition, 134*, 110–120. http://dx.doi.org/10.1016/j.cognition.2014.09.001.

Raven, J. C. (1941). Standardization of progressive matrices, 1938. *The British Journal of Medical Psychology, 19*, 137–150. http://dx.doi.org/10.1111/j.2044-8341.1941.tb00316.x.

Resnick, I., Davatzes, A., Newcombe, N. S., & Shipley, T. F. (2016). Using relational reasoning to learn about scientific phenomena at unfamiliar scales. *Educational Psychology Review*. http://dx.doi.org/10.1007/s10648-016-9371-5.

Richland, L. E., Begolli, K. N., Simms, N., Frausel, R. R., & Lyons, E. A. (2016). Supporting mathematical discussions: The roles of comparison and cognitive load. *Educational Psychology Review*. http://dx.doi.org/10.1007/s10648-016-9382-2.

Rindermann, H., & Neubauer, A. C. (2004). Processing speed, intelligence, creativity, and school performance: Testing of causal hypotheses using structural equation models. *Intelligence, 32*(6), 573–589. http://dx.doi.org/10.1016/j.intell.2004.06.005.

Runco, M. A., & Albert, R. S. (1986). The threshold theory regarding creativity and intelligence: An empirical test with gifted and nongifted children. *Creative Child & Adult Quarterly, 11*(4), 212–218.

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*(4), 361–368. http://dx.doi.org/10.1080/10400419.2010.523393.

Runco, M. A., Noble, E. P., Reiter-Palmon, R., Acar, S., Ritchie, T., & Yurkovich, J. M. (2011). The genetic basis of creativity and ideational fluency. *Creativity Research Journal, 23*(4), 376–380. http://dx.doi.org/10.1080/10400419.2011.621859.

Sala, G., & Gobet, F. (2017). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology, 53*(4), 671–685. http://dx.doi.org/10.1037/dev0000265.

Shi, B., Wang, L., Yang, J., Zhang, M., & Xu, L. (2017). Relationship between divergent thinking and intelligence: An empirical study of the threshold hypothesis with Chinese children. *Frontiers in Psychology, 8*. http://dx.doi.org/10.3389/fpsyg.2017.00254.

Silvia, P. J. (2008). Another look at creativity and intelligence: Exploring higher-order models and probable confounds. *Personality and Individual Differences, 44*(4), 1012–1021. http://dx.doi.org/10.1016/j.paid.2007.10.027.

Silvia, P. J. (2011). Subjective scoring of divergent thinking: Examining the reliability of unusual uses, instances, and consequences tasks. *Thinking Skills and Creativity, 6*(1), 24–30. http://dx.doi.org/10.1016/j.tsc.2010.06.001.

Silvia, P. J. (2015). Intelligence and creativity are pretty similar after all. *Educational Psychology Review, 27*(4), 599–606. http://dx.doi.org/10.1007/s10648-015-9299-1.

Silvia, P. J., Christensen, A. P., & Cotter, K. N. (2016). Commentary: The development of creativity—Ability, motivation, and potential. In B. Barbot (Vol. Ed.), *Perspectives on creativity development: Vol. 151*, (pp. 111–119). San Francisco, CA, US: Jossey-Bass.

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., ... 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. http://dx.doi.org/10.1037/1931-3896.2.2.68.

Sligh, A. C., Conners, F. A., & Roskos-Ewoldsen, B. (2005). Relation of creativity to fluid and crystallized intelligence. *The Journal of Creative Behavior, 39*(2), 123–136. http://dx.doi.org/10.1002/j.2162-6057.2005.tb01254.x.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., ... Dougherty, M. R. (2013). Training working memory: Limits of transfer. *Intelligence, 41*(5), 638–663. http://dx.doi.org/10.1016/j.intell.2013.07.013.

StataCorp (2013). *Stata statistical software: Release 13.* College Station, TX: StataCorp LP.

Sternberg, R. J. (2006a). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence, 34*(4), 321–350. http://dx.doi.org/10.1016/j.intell.2006.01.002.

Sternberg, R. J. (2006b). The nature of creativity. *Creativity Research Journal, 18*(1), 87–98. http://dx.doi.org/10.1207/s15326934crj1801_10.

Sternberg, R. J., Grigorenko, E. L., Ngorosho, D., Tantufuye, E., Mbise, A., Nokes, C., ... Bundy, D. A. (2002). Assessing intellectual potential in rural Tanzanian school children. *Intelligence, 30*, 141–162.

Sternberg, R. J., & Lubart, T. I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity.* New York, NY: Free Press.

Tarr, G. (2012). Small sample performance of quantile regression confidence intervals. *Journal of Statistical Computation and Simulation, 82*(1), 81–94. http://dx.doi.org/10.1080/00949655.2010.527844.

Torrance, E. P. (1962). *Guiding creative talent.* Englewood Cliffs, NJ, US: Prentice-Hall Inc.

Torrance, E. P. (1972). Predictive validity of the Torrance Tests of Creative Thinking. *The Journal of Creative Behavior, 6*(4), 236–252.

Trey, L., & Khan, S. (2008). How science students can learn about unobservable phenomena using computer-based analogies. *Computers and Education, 51*(2), 519–529. http://dx.doi.org/10.1016/j.compedu.2007.05.019.

Turner, M. A. (1999). Generating novel ideas: Fluency performance in high-functioning and learning disabled individuals with autism. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 40*(2), 189.

van de Kamp, M., Admiraal, W., van Drie, J., & Rijlaarsdam, G. (2015). Enhancing divergent thinking in visual arts education: Effects of explicit instruction of metacognition. *British Journal of Educational Psychology, 85*(1), 47–58. http://dx.doi.org/10.1111/bjep.12061.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*(3), 289–335. http://dx.doi.org/10.1016/j.jml.2003.10.003.

Vock, M., & Holling, H. (2008). The measurement of visuo-spatial and verbal-numerical working memory: Development of IRT-based scales. *Intelligence, 36*(2), 161–182. http://dx.doi.org/10.1016/j.intell.2007.02.004.

Wilson, A., Reich, B. J., Nolte, C. G., Spero, T. L., Hubbell, B., & Rappold, A. G. (2017). Climate change impacts on projections of excess mortality at 2030 using spatially varying ozone–temperature risk surfaces. *Journal of Exposure Science and Environmental Epidemiology, 27*(1), 118–124. http://dx.doi.org/10.1038/jes.2016.14.

Yang, J. S., Hansen, M., & Cai, L. (2012). Characterizing sources of uncertainty in item response theory scale scores. *Educational and Psychological Measurement, 72*(2), 264–290. http://dx.doi.org/10.1177/0013164411410056.

Yi, X., Hu, W., Plucker, J. A., & McWilliams, J. (2013). Is there a developmental slump in creativity in China? The relationship between organizational climate and creativity development in Chinese adolescents. *The Journal of Creative Behavior, 47*(1), 22–40. http://dx.doi.org/10.1002/jocb.21.

Yu, K., Lu, Z., & Stander, J. (2003). Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician), 52*(3), 331–350. http://dx.doi.org/10.1111/1467-9884.00363.

Zietz, J., Zietz, E. N., & Sirmans, G. S. (2008). Determinants of house prices: A quantile regression approach. *The Journal of Real Estate Finance and Economics, 37*(4), 317–333. http://dx.doi.org/10.1007/s11146-007-9053-7.