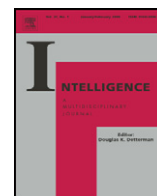




Contents lists available at SciVerse ScienceDirect

Intelligence

journal homepage:



## Creative reasoning across developmental levels: Convergence and divergence in problem creation

Saskia Jaarsveld<sup>a</sup>, Thomas Lachmann<sup>a,\*</sup>, Cees van Leeuwen<sup>b,c</sup>

<sup>a</sup> Department of Psychology II, University of Kaiserslautern, Kaiserslautern, Germany

<sup>b</sup> Laboratory for Perceptual Dynamics, Brain Science Institute RIKEN, Wako-shi, Saitama, Japan

<sup>c</sup> University of Leuven, Belgium

### ARTICLE INFO

#### Article history:

Received 18 November 2010

Received in revised form 2 December 2011

Accepted 3 January 2012

Available online xxxxx

#### Keywords:

Creative reasoning

Creativity

Intelligence

Development

Convergent thinking

Divergent thinking

Problem solving

### ABSTRACT

We recently proposed the Creative Reasoning Test (CRT), a test for reasoning in ill-defined problem spaces. The test asks children who first performed the Standard Progressive Matrices test (SPM) to next generate an SPM-style test item themselves. The item is scored based on different aspects of its complexity. Here we introduce a method to provide separate sub scores for convergent and divergent production. A sample of 205 children from Grades 1 to 4 performed in both SPM and CRT, followed by a standard creativity test, the Test of Creative Thinking–Drawing Production (TCT–DP). We considered developmental aspects of the CRT scores as well as how they relate to classical reasoning and creativity test scores. Whereas the convergent production sub score of the CRT correlated with the SPM, the divergent production sub score correlated with the TCT–DP and no correlation was observed between both sub scores of the CRT. We concluded that convergent and divergent abilities in creative production can be assessed independently, within the same knowledge domain as classical problem solving.

© 2012 Elsevier Inc. All rights reserved.

### 1. Introduction

The relationship between intelligence and creativity is a topic of longstanding controversy (Plucker & Renzulli, 1999; Runco, 2007). Correlations between intelligence and creativity test scores vary widely, depending on how these constructs are measured. We introduced a creative reasoning test (CRT), in which both constructs are measured using the same task (Jaarsveld, Lachmann, Hamel, & van Leeuwen, 2010). Here, we introduce a refined scoring technique for the CRT and investigate how our measurements relate to those of classical intelligence and creativity test scores across developmental levels.

Whereas some authors failed to spot a link between intelligence and creativity (Eisenman, Platt, & Darbes, 1968; Kim,

2005; MacKinnon, 1962; Sen, 1993; Smilansky, 1984; Wallach & Kogan, 1965), others found a positive linear relationship (Barron & Harrington, 1981; Guilford, 1967; Preckel, Holling, & Wiese, 2006; Runco & Albert, 1986; Silvia, 2008a; Yamamoto, 1964). Still others found the relation to disappear above a certain threshold level of IQ (e.g. IQ > 120; Guilford, 1967; Jung et al., 2009). According to Sligh, Conners, and Roskos-Ewoldsen (2005), these comparisons suffer from a common methodological flaw: variance in IQ is smaller in the high IQ than the average IQ group. Controlling the variance resulted in a positive relation for all levels of intelligence. Distinguishing between subtypes of general intelligence known as fluent and crystallized intelligence (Cattell, 1967), Sligh et al. (2005) observed an *inverse* threshold effect with fluid IQ: a correlation with creativity test scores in the high IQ group but not in the average IQ group. Intelligence defined as fluid IQ, verbal fluency, and strategic abilities shows a higher correlation with creativity scores (Silvia, 2008b) than when defined as crystallized intelligence. Finally, creativity might have a stronger relationship

\* Corresponding author at: Faculty of Social Sciences, Department of Psychology II, University of Kaiserslautern, Edwin-Schroedinger-Strasse 57, D-67663 Kaiserslautern, Germany.

E-mail address: lachmann@rhrk.uni-kl.de (T. Lachmann).

with *divergent* than with *convergent* thinking (Guilford, 1956, 1967). Yet, creativity tests which involve convergent thinking (e.g., Remote Association Test; Mednick, 1962), show higher correlations with intelligence than ones that involve only divergent thinking (e.g., the Alternate Uses Test; Guilford, Christensen, Merrifield, & Wilson, 1978).

### 1.1. Creativity

The link between intelligence and creativity remains elusive; perhaps this is because the concept of creativity has not solidified. Concepts of creativity vary in the degree to which they take on board elements of what is generally considered to be intelligence. In a classical perspective, creativity is defined as the ability to innovate and move beyond what is already known (Ghiselin, 1952/1985; Vernon, 1970; Wertheimer, 1945/1968). In other words, it emphasizes the aspect of innovation. This involves the ability to consider things from an uncommon perspective, transcend the old order (Chi, 1997; Ghiselin, 1952/1985; Ward, 2007), and explore loosely associated ideas (Boden, 1990; Christensen, 2007; Gentner, 1983; Guilford, 1950; Koestler, 1964; Mednick, 1962). These abilities may to various degrees be related to divergent thinking. This construct is measured typically by standard creativity tests, such as the Bennett (1944), and the Torrance (1966). These include subtests such as: Asking Questions, Product Improvement, and Unusual Uses. These tests ask for the generation of as many ideas as possible and they score fluency, flexibility and originality of generated ideas (as defined in Guilford, 1967).

Creativity could also be defined as the ability to generate a solution to an ill-defined problem (Getzels & Csikszentmihalyi, 1976; Wertheimer, 1945/1968). Ill-defined problems have constraints that are inexplicit about how the solution should fit the criteria (Akin & Akin, 1998). To arrive at a solution, divergent thinking alone is not enough; ideas need to be forged into one clear and appropriate proposal (Bear, 2003; Jaarsveld & van Leeuwen, 2005), using models of the design's anticipated functional characteristics (Elias & Dasgupta, 2005; Viswanathan & Linsey, 2009). In the Geneplore model (Finke, Ward, & Smith, 1992) functional characteristics do constrain the generation or exploration phases of the process. The generation phase brings forth pre inventive objects, imaginary objects that are generated without any constraints in mind. In exploration, these objects are assessed for their possible functionalities. The creative process is not so much guided by constraints as by associations: What could be the use of this imaginary object? In anticipating the functional characteristics of generated ideas, convergent thinking (Guilford, 1967) is needed to apprehend the situation, make evaluations (Kozbelt, 2006, 2008), and consider the consequences of a chosen solution (Goel & Pirolli, 1992). Convergent reasoning in creativity tasks invokes criteria of functionality and appropriateness (Halpern, 2003; Kaufmann, 2003), goal directedness and adaptive behavior (Sternberg, 1982b), as well as the abilities of planning and attention; in other words, abilities which typically play a role in intelligent behavior (Dehn & Schank, 1982; Shavinina, 2001). Convergent thinking phases may even require divergent thinking subprocesses to identify restrictions on proposed new ideas and suggest requisite

revision strategies (Mumford, Hunter, Eubanks, Bedell, & Murphy, 2007).

Furthermore, creativity can be defined as the ability to identify problematic aspects of a given situation (Ghiselin, 1952/1985) and, in a wider sense, as the ability to define completely new problems (Getzels, 1975, 1987; Runco, 1990). Dehn and Schank (1982) emphasized that in defining new problems, the organization of knowledge in memory plays an important role. These authors state that the adaptability of this organization is important for generating quality solutions and, therefore, creativity could be defined as paramount to intelligence (Haensly & Reynolds, 1989). As far back as 1950, Guilford stated that creativity extends well beyond intelligence (Guilford, 1987). Creativity is as much about what we explicitly know as about what we can do with it, i.e., as much about knowledge apprehension as about knowledge application (Basadur, 2005).

### 1.2. Intelligence

Concepts of intelligence vary, likewise, in the degree to which they take on board elements of what is generally considered to be creativity. The traditional view defines intelligence as the ability to find the correct solution to well-defined problems, such as a proof to a theorem (Simon, 1973). Problem-solving tasks have definite goals and criteria to evaluate solutions, which are found through specifiable methods or algorithms. Well-defined problems are typically given in standard intelligence tests (e.g., Cattell, 1949; Miller, 1970, and Raven, 1938/1998). Solving these problems involves the cognitive abilities of convergent thinking (Guilford, 1959) and reasoning. Reasoning may be characterized as an attempt to combine elements of old information to form new information (Sternberg, 1982a). Reasoning abilities depend on the organization of knowledge in the relevant domain (Mumford, Baughman, & Sager, 2003).

Intelligence may also be seen as a domain-general ability (*g*-factor; Cattell, 1967; Spearman, 1904) that has much in common with metacognitive functions (Campion, Brown, & Ferrara, 1982), such as comprehension, invention, direction, and criticism (Binet, 1911). We may understand intelligence in terms of problem solving ability (Simon & Newell, 1971). Problem solving can then be approached in terms of rules operating on problem spaces. These rules are active transformations that may bring a problem closer to a solution. Here, the issue of the relationship of intelligence and creativity returns in the form of the question: what is the relation between *classical* and *creative* problem solving. In classical problem solving, the rules that may lead to the solution operate in a well-defined problem space and are understood as given: they are either explicitly defined as the rules of the game or tacitly assumed to be specified by the problem situation. Creative problem solving appears to involve rules that have to be invented. Creative problem solving breaks the implicitly defined rules of the standard problem, it operates in an ill-defined problem space. This may require a specific form of reasoning, i.e. creative reasoning.

Whereas the classical problem solving approach requires steps and goals to be prespecified at all stages, creative reasoning requires an open-ended, adaptive and flexible attitude towards emerging situations. This extended concept of

intelligence, it appears, shares functional components with creativity, especially one aspect of creative abilities, i.e. flexibility. Niaz and De Nunez (1991) showed that high-flexible participants scored best on the Torrance Test of Creative Thinking, whereas low-flexible ones performed better on a formal reasoning test.

An even more inclusive view to intelligence considers flexibility to adapt to the idiosyncrasies of the moment as the ultimate manifestation of cognition (Calvin, 1998; Colunga & Smith, 2008; Wechsler, 1981). This perspective reflects the importance of flexibility, where an idea revolves around an ill-defined set of constraints, while dealing with multiple idiosyncrasies of the moment (Jaarsveld & van Leeuwen, 2005). According to Sternberg (2005) achieving this level of flexibility requires a synthesis of wisdom, intelligence, and creativity. Even though these three are still considered as referring to different entities (Bear, 2003; Sternberg & Lubart, 2003), some models have proposed that both types of thinking abilities are linked in each stage of the creative process (Amabile, 1990; Liu, Chakrabarti, & Bligh, 2002; Neubauer, 2010; Parnes, 1972). If intelligence is linked inextricably with creativity (Fuster, 2005), we may raise the question if one, integral concept might be preferable (Kaufmann, 2003). Or, alternatively, we may wish to maintain that they are independent abilities, which interact in dynamic and intricate ways to flexibly solve problems (Jaarsveld & van Leeuwen, 2005).

### 1.3. Creative reasoning

The literature is moving towards the view that creativity and intelligence play complementary roles (Runco, 2003; Szobiová, 2001; Ward, 2007). A previous study using a creative design task illustrates this (Jaarsveld & van Leeuwen, 2005). Non expert participants in a design process made a series of sketches. These sketches provided evidence of divergent thinking for generating new approaches and ideas, as well as of convergent thinking for prioritizing these ideas in light of the requirements of the solution. We observed that these processes were taking turns in the creative process, and that in the early stages of the design, divergent activity predominated, whereas convergent thinking took the lead in the later stages. It was concluded that divergent production, typically seen as part of a creative process, and convergent reasoning, as traditionally associated with intelligence, show a delicate interplay throughout the design process.

There is little consensus about terminology when intelligence and creativity engage in complementary roles. The term: “creative interaction” (Necka, 2003) expresses the idea that creation and evaluation are both aspects of the same mental operation, and not two distinct stages. “Evaluative thinking” (Bear, 2003) denotes a broad class of loosely related cognitive skills that operate on generated ideas. Even more broadly, “critical thinking” (Halpern, 2003) denotes the application of any cognitive skills or strategies that increase the probability of a desirable outcome. Ultimately, the broader the concepts become, the less meaningful they are. We proposed the term *creative reasoning* (Jaarsveld et al., 2010) to denote the specific processes in which divergent and convergent abilities play complementary roles in a creative problem solving task. The term emphasizes the mutually supportive contributions of and cooperation between

both types of thinking in processes that evolve from ill-defined problem situations and demand original and appropriate solutions.

If divergent and convergent operations work together in cognitive reasoning tasks, the question can still be raised, whether they are correlated or independent abilities. To answer this question, it is common to perform a test battery on the same group of participants, consider the correlation matrix of the test items, submit the results to factor analysis (Velicer & Jackson, 1990) or latent variable models (Unsworth, Redick, Lakey, & Young, 2010) and see if the outcome is a one or two-factor solution. However, the domains of knowledge may differ strongly between these tasks—and this could depress the correlations (a problem discussed at length by Jaarsveld et al., 2010). To minimize this problem, a single domain of knowledge would be preferable.

To satisfy this requirement we will consider a problem finding task (Smilansky, 1984; Smilansky & Halberstadt, 1986). These authors asked individuals first to perform the last two sets, Sets D and E, of the Standard Progressive Matrices test, (SPM, Raven, 1938/1998), and afterwards to create an SPM test item, i.e. a Raven matrix, of their own. Using this task secured that the knowledge domain in the well and the ill-defined task was the same. Both types of tasks require the understanding of a set of relations and reasoning about possible solutions. To evaluate the items invented, Smilansky scored for the number of what he called elements (which are, unfortunately, not defined in detail) and the relations among these elements. He correlated this measure with the individual scores of the partially performed Raven test. One of his results was that, within a group of students, the combination of low RPM scores and high scores in the generation task failed to occur. From this, it could be concluded that creativity builds on established intelligence.

The problem finding task initiated by Smilansky (1984) may be considered as a test of creative reasoning (CRT). It both requires divergent production, as test items need to be generated and convergent operations, as the items generated need to fulfill the kind of rules that enable matrix completion in the SPM. In fact, the overlap in knowledge domains between CRT and SPM facilitates the comparison of creative reasoning and classical intelligence. Standard creativity and intelligence tests operate on different knowledge domains, and therefore run the risk of confounding the comparison with domain-specific functions. Jaarsveld et al. (2010) compared reasoning in the CRT and the SPM and found that, despite the identity of the problem domain the operations used differ between the tasks as a function of the differences in problem spaces. In this study, the reasoning operations used in SPM and CRT were described and compared. The major finding of this study was that despite the fact that the SPM immediately preceded the CRT; the frequency with which different categories of reasoning operations were applied in the CRT was independent of that in the SPM (Jaarsveld et al., 2010).

To observe whether convergent and divergent abilities are dependent or independent, we need a method for scoring both separately in the CRT. Jaarsveld et al. (2010) provided a scoring method for the CRT optimized for comparing it with the SPM. The method uses the same scoring rules for problem creating and problem solving, allowing a direct quantitative

comparison of the results. The rules are based on a review of the kind of relations between components that could occur in Raven items. The relations are typically transformations from one component of the matrix to another. In the SPM, these transformations are applied across rows or columns of the matrix in order to identify a missing component from amongst a set of alternatives. Whereas only a limited set of transformations occur in the SPM, in the CRT, people could, in principle, apply any transformation for creating an item. Jaarsveld et al. (2010) aimed for a complete listing and valuation of any of the transformations between components that led to the completion of an item. With this method, we were able to answer the question, whether individuals who have just solved SPM items in which certain transformations were featured, apply these same transformations when they design a new matrix. We found that relations featuring in the solving task differed strongly from those applied in the problem creation task. It was concluded that creative reasoning, as measured by the CRT, does not reflect SPM solving ability, and that both cognitive abilities develop rather independently from each other from Kindergarten to secondary school.

Here, we elaborated on the scoring method of the CRT, in order to obtain separate sub scores of convergent and divergent thinking. CRT items by different individuals will be scored for both abilities, and scores will be compared within individuals. Because they are scored for one and the same creative reasoning task, presence or absence of a correlation can be interpreted as reflecting the dependence or independence, respectively, of convergent and divergent abilities. To investigate whether these sub scores hold relevance in a broader framework we will compare them with standard problem solving and creativity test scores. We predict that convergent thinking sub scores will relate to problem solving test scores and divergent thinking sub scores will relate to standard creativity test scores. As in Jaarsveld et al. (2010) we also studied the developmental aspects of both cognitive abilities.

## 2. Method

### 2.1. Participants

Participants were native German speaking children from a primary school in Kaiserslautern, Germany (Pestalozzi Grundschule). We tested girls ( $N=101$ ) and boys ( $N=104$ ) from Grade 1 to Grade 4, aged between seven and ten years. Descriptive statistics for each grade are shown in Table 1. The state authorities had given their approval to

**Table 1**  
Numbers of participants ( $N=205$ ) by grade and their age in month.

Grade	N	%Girls	Age in month			
			Min	Max	M	SD
1	51	54.9	70	93	84.67	4.642
2	43	41.9	87	109	97.86	4.617
3	51	51	100	124	108.82	5.428
4	60	48.3	111	129	120.65	4.129

Note. Percentage of girls shows that there was not a large difference between the number of girls and boys.

this study, and the parents were asked to sign a letter of consent. Parents and children were informed that no results would be reported back to the school. Each class received a shopping voucher and the children received toys and candies in reward for their participation. We excluded participants from analysis whose performance on the CRT was deemed invalid ( $n=5$ ), because there was evidence that the instructions were not taken seriously. Two of the excluded children were in Grade 2, two were in Grade 3, and one in Grade 4. We further excluded participants whose SPM results were considered invalid; the SPM manual states that “for general purposes the total score appears to be relatively valid even when discrepancies of more than two points” occur in a series (p. 59). This led us to exclude participants who showed two or more discrepancies above 3 points. This decision resulted in exclusion of 27 children: 16 from Grade 1, five from Grade 2, two from Grade 3, and four from Grade 4. The remaining sample consisted of 205 children in total.

### 2.2. Material

The German versions of three paper-and-pencil tests were applied in the following order for each participant: 1. The Standard Progressive Matrices-classic version (SPM-C; Raven, 1938/1998), 2. The Creative Reasoning Task (CRT; Jaarsveld, S., in prep.), and 3. The Test for Creative Thinking–Drawing Production (TCT–DP, Test zum Schöpferisches Denken–Zeichnerisch, Urban & Jellen, 1995).

#### 2.2.1. SPM-C

The SPM-C is a non-verbal test that can either be presented to participants without a strict time limit in order to “assess their capacity for coherent and clear thinking” (power test) or with a time limit “in order to assess their intellectual efficiency when working under stress” (Raven, Raven, & Court, 1998, pp. 2 and 6). We used the power test version of the SPM-C. The test was designed for children and adults and is intended to be free of education and social background influences. Correlations of the SPM-C with the Stanford–Binet test and WAIS (Wechsler Adults Intelligence Scale, Wechsler, 1981) showed values of  $r=.54$  for children and  $r=.86$  for teenagers (criteria-related validity). Correlations of the SPM-C with next year's school performance (predictive validity) for English and German speaking children and adults showed values around  $r=.70$  (Horn, 2009). Test–retest reliabilities for the SPM-C showed values of  $r=.84$  after one year (high school students) and of  $r=.61$  after four years (12–21 year old individuals, cf., Standard progressiver matrices, manual, Kratzmeier & Horn, 1988). Carpenter, Just, and Shell (1990) found a correlation between error rates on the SPM-C (Sets I and II) and the Tower of Hanoi task of  $r=-.77$  ( $N=43$ ) because both tasks rely on generating and maintaining goals and sub goals in working memory. Marshalek, Lohman, and Snow (1983) showed that the correlations between  $g$ -factor and other scores of psychometric tests increase with the apparent complexity of the required cognitive operations. Complex tests were defined as having more than 50% of their variance accounted for by  $g$  (with a maximum of 60%), while simple tests had less than 25% of their variance accounted for. The authors

showed that the Raven test (Advanced Progressive Matrices, Raven, 1976), requiring abstract problem solving, analysis, and inference of rules, hence a complex test, had 60% of its variance accounted for by *g* and showed a correlation of  $r = .66$  with the WAIS.

The items of SPM-C test are presented in a booklet containing five series (Series A, B, C, D, E) of twelve items each. They have several *formats* as illustrated in Fig. 1: a continuous pattern in a large frame with the lower right part missing for completion (Fig. 1A), or a pattern divided into a 2×2 (Fig. 1C) or a 3×3 matrix (Fig. 1D). The pattern determines the relations between its components, such as: sameness (# # #), increase (I II III), or combination (∖/X). Note that a pattern may contain several relations at a time, for instance sameness in the horizontal and increase in the vertical direction. In the combination example, with ∖=A and /=B, the pattern reads as A B AB, showing a sequence of three components, the last one with two subcomponents. In principle, any piece of a graphic can be a component or subcomponent, depending on the relation in which they feature, as long as it is relevant to the solution. Each cell of the matrix contains at most one component of the pattern. In the matrix the lower right cell remains empty for completion. A set of six cells (Series A and B) or eight cells (Series C to E) is presented as response alternatives below the matrix, from which the correct completion has to be chosen and crossed out on an answering sheet.

In the standard procedure for scoring the test, correctly solved items are assigned the value one and summed into sub totals which indicate the number of items solved correctly per series. The sum of the sub totals represents the score of the SPM-C. Because the series become increasingly difficult to solve the sub totals are expected to decrease and they can be used to assess the consistency of the work produced (Raven et al., 1998, pp. 59 and 69). In other words, the number of items solved correctly in the easier series predicts the number of items that should be solved correctly in the more difficult series. When this expectancy is not matched it is assumed that an individual's performance was inconsistent and that the current score does not reflect the individual's reasoning abilities. For each SPM-C score the manual gives a sequence of sub totals and allows for a discrepancy of plus two points to minus one point.

2.2.2. CRT

The CRT asks participants to *create* a Raven item in a frame, which is printed on the test form. Participants can chose to work in this frame or to draw lines indicating a matrix. They are instructed that the item should be new, as complex as possible and solvable (“make your item as hard as possible to solve for someone else”). In contrast to the preceding study (Jaarsveld et al., 2010) we did not ask participants to create multiple alternative solutions in special boxes below the matrix; they only had to draw the correct solution inside the matrix itself, i.e. they drew a completed matrix.

Creating an item involves generating components as well their relations. Accordingly, the CRT should contain two sub scores: One for scoring convergent thinking abilities in an ill defined problem space and a second one for scoring divergent thinking abilities in the same space. Here we will introduce these sub scores. The first sub score, *Relations*, scores the extent in which a participant has succeeded in expressing relations between components or their subcomponents (for instance, Change, Combination and Succession), the second one, *Components & Specifications*, scores the number of components and specifications created (see Jaarsveld et al., 2010). Components and subcomponents are identified by the rules in which they occur. Each can, in principle, occur in more than one rule. Specifications are particular textures and alterations, such as zoom or rotate which are applied to components and are ornamental, i.e. irrelevant to the relations.

A number of modifications were made from the original version of the CRT (Jaarsveld et al., 2010). The present version valuates not only the number of relations but also their complexity and whether a relation was applied over all rows and columns. This method produces a larger range of score values and therefore provides more differentiated information of reasoning abilities. In addition, relations are now labeled according to the *format* in which they are expressed. Drawings could have one of the following formats: Continuous Pattern, String, Matrix-2×2, and Matrix-3×3 (see Fig. 1).

2.2.2.1. *Relations*. The relations and the scores that their applications in the CRT receive are listed in Table 2. Since the CRT problem is open, by definition it is impossible to fix a set of categories for all possible solutions. However, the current

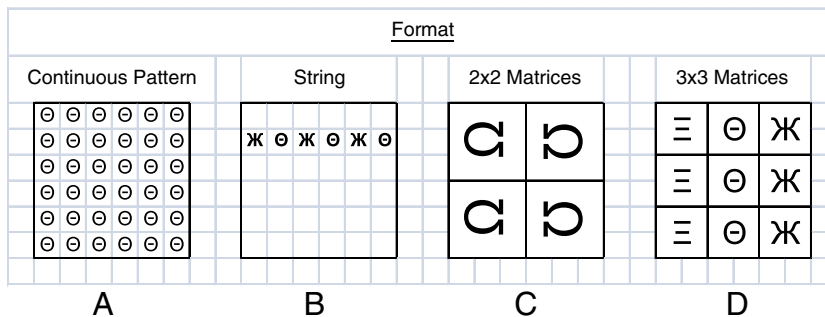


Fig. 1. Formats of matrices for SPM-C and CRT with examples of Relations discussed in the text. A: the format Continuous Pattern with Relation Pattern Completion; B: the format String with the relation Iteration of two or >2 Components—this format does not feature in the SPM-C but occurs in the CRT; C: the format 2×2 Matrices with the relation Symmetry; D: the format 3×3 Matrices with the relation Change.

**Table 2**  
Descriptive statistics of the relations per format used to score matrices in SPM and CRT.

Relation	Actual		Observed							
	Min	Max	n	Min	Max	M	SD	Skewness	Kurtosis	
Continuous pattern										
1	Ideosyncratic coherence	-	1	39	1	1	1.00	0.00	-	-
2	Jigsaw	-	2	2	2	2	2.00	0.00	-	-
3	Pattern completion	-	3	68	3	3	3.00	0.00	-	-
String										
4	Iteration of one component	2	9	8	2	7	4.13	1.73	0.63	-0.8
5	Iteration of two or >2 components	6	19	10	6	13	8.90	2.60	0.52	-1.31
Matrix 2x2										
6	Symmetry	3	8	3	3	8	5.33	2.52	0.59	-
7	Change and increase	6	19	4	6	17	9.50	5.07	1.85	3.49
8	Succession 2x2	7	10	2	10	10	10.00	0.00	-	-
Matrix 3x3										
9	Change	8	27	11	8	27	19.91	7.05	-0.57	-1.28
10	Increase	9	52	9	9	39	23.56	10.51	-0.09	-0.80
11	Succession 3x3	10	36	15	10	36	23.67	10.67	-0.03	-1.77
12	Combination and indication of mathematical operation	6	96	12	7	40	24.83	11.31	-0.32	-0.92
13	Two values	12	128	0	-	-	-	-	-	-
14	Contrast	8	39	0	-	-	-	-	-	-
15	Groups of components	3	13	15	3	13	10.80	3.03	-	1.64
Entries for scored relations <sup>a</sup>			198							
Participants who received no score			9							
			Subtotal							
			207							
Participants who applied two different relations			4							
			Participants total							
			205							

Note. Minimum scores represent the valuation of a matrix that is incomplete and meets the minimum conditions for assigning a specific relation to the matrix. Maximum scores are the scores for applying the most complex variation of a relation in a complete matrix. A CRT matrix can contain several relations. For the relations: Idiosyncratic Coherence, Jigsaw, and Pattern Completion there is only a unique score.

<sup>a</sup> Participants who applied the same relation more than once (n = 11) were entered with the application that scored most.

set covers all the ones actually occurring. It is originally based on a study by Carpenter et al. (1990). The modifications made with respect to these rules are described in Jaarsveld et al. (2010). Two major improvements with respect to this study are the format String, which accommodates generated relations contained in strings of components (Fig. 1), and a new method for scoring the Relations, which now allows for scoring incomplete matrices. Other, minor modifications in reference to the present study are mentioned in the following. Table 2 also lists the minimum and maximum score that can be attained per relation. The valuation reasonably reflects the complexity of the rule, as it reflects both the scope (number of components covered) and the number of transformations on relative positions or types of components. Because, in principle, an unlimited number of rules can be combined in a single generation, the number of points that a single item can score theoretically have no upper limit.

The format *Continuous Pattern* contains the relations (1) *Idiosyncratic and Semantic Coherence*, which is simply a seemingly arbitrary collection of components; the choice of the completion is transparent, if at all, only to the maker; (2) *Jigsaw*, which applies to drawings divided into jigsaw pieces; and (3) *Pattern Completion*, which applies when a repetitive motive is continued in the completion. In Jaarsveld et al. (2010), none of the participants applied Jigsaw in the CRT, so this is a novel addition to the categories of relations.

The new format *String* refers to a single row or column of components. It contains the relations (4) *Iteration of one Component*, as in the string *A A A...*, (5) *Iteration of a chunk*

of two or more components, as in *C D E C D E*. The format *Matrix-2x2* contains the relations (6) *Symmetry*, (7) *Change & Increase*, and (8) *Succession*. The relation *Symmetry* refers to the spatial mirror image transformation; *Change* applies when components are same in a row but different in a column or vice versa (e.g., first row *A A* and *B B* in the second); *Increase* shows components as increasing or decreasing in number or size. Because in  $2 \times 2$  matrices the *Increase* relation is indistinguishable from *Change*, they form a single category for these matrices and receive identical valuation. *Succession* applies when an identical cycle of components appears between rows or columns, but with a shift between components from one row or column to the next (e.g., a first row *\$ \$* and a second row *\$ \$*). The format *Matrix-3x3* has (9) *Change*, (10) *Increase*, (11) *Succession*, (12) *Combination & Indication of Mathematical Operation*, (13) *Two Values*, (14) *Contrast*, and (15) *Group of Components*. *Change* and *Increase* are defined as previous, but now form separate categories because they are distinguishable for  $3 \times 3$  Matrices. Also, they receive different valuations, based on the scoring method used (see below). The difference in valuations is consistent with the observation that *Increase* features in later, more difficult items of the SPM-C than *Change*. *Combination* applies when components in the two cells of a row or column are combined to form a new component in a third cell (e.g., *A* and *B* form *AB*). *Indication of Mathematical Operation* occurs when two items from a row or column are combined into a third component by an operation, in which a graphical representation of numerical operation can be recognized. The

graphical characteristic of the second component indicates what type of mathematical operation must be applied on the first in order to get the third. For instance: first row II III IIII, indicating  $2 + 3 = 5$  and second row II III I, indicating  $2 - 3 = -1$ . Two Values applies when three components each feature twice over three cells of a row or column (AB BC AC, e.g. SPM-C Item E11).

Contrast (equals *Remain* in Jaarsveld et al., 2010) applies when a subcomponent which is shared by two components features in the third while others that are not shared disappear. As a consequence, the shared subcomponent features in all three components and the ones that are not shared feature only once in a row or column (e.g., SPM-C Items E8 and E9). The subcomponents which are not shared are of the same type, e.g. both are half circles, but they differ in their orientation. They even can be of the same type as the shared ones. The rule Contrast features in the SPM together with Combination or Two Values. Applied on its own the former should not be a difficult rule to solve; however, it is difficult to discern when applied in a matrix together with the other rules.

Group of Components applies when the distribution of a component across the matrix has a 2-dimensional regularity. For instance, the same component occurs in the center cell and in all four corners of the matrix, together forming a five of dice.

In contrast to Jaarsveld et al. (2010) where scoring was applied only to complete and correctly generated matrices, with the present scoring method, for  $3 \times 3$  matrices, we are able to score rules per row or column, even if the matrix as a whole is not correct. This allows a much more detailed evaluation of the thinking process and a better differentiation between individuals. For instance, Indication of Mathematical Operation can apply to a single row or column. The same is true for the relations Combination, Two Values, and Contrast. The relations Change and Succession, in contrast, apply across rows or columns, and thus the minimal requirement is that the rule applies to a row or column and one element in an adjacent row or column. Thus, for a relation to occur, a minimal set of conditions has to be fulfilled. As a result, scoring ambiguities of incomplete matrices cannot arise.

The new scoring of Relations depends on the format of the items. For matrices in the Continuous Pattern format, Idiosyncratic scores 1, Jigsaw Puzzle 2, and Texture 3. This ordering is based on the increasingly constrained character of the pattern completion. For the format String, Iteration of one Component is simpler than Iteration of two or more Components, and the latter therefore receives a higher valuation. For  $2 \times 2$  and  $3 \times 3$  Matrices, scoring is done in several steps. First, we listed the relations that apply to the item and for each relation marked the (sub) components it covered. Next, for each relation we assigned an index value  $i = 1, 2, \dots$  to all first appearances of the marked (sub)components, starting from the top-left cell of the item, proceeding from left to right through each row from top to bottom. Third, passing through the matrix in the same order as previous, we accumulate a score, in which the first encounter of a (sub) component is scored with a value identical to its index; each next time we encounter a (sub) component again, we assign the same score as previously, incremented with 1 when it

occurs in a row different from where it has previously been encountered, and with another 1 when it occurs in a column in which it has not previously been encountered. The resulting score is the sum total of all values assigned to (sub) components of the matrix.

To optimize the differentiation between items, some restrictions and extensions apply to this scoring method. For strings, Iterations of one component have a maximum score of 9, for Iterations of a chunk of two or more components the maximum score includes three chunks, i.e., two iterations. A final refinement is that in  $3 \times 3$  matrices the directional aspect of the relations Increase and Combination is taken into account. These relations are given extra points, depending on the location of the third component of the relation (Right, Bottom, Left, Top, and Center). This was done, based on the observation that items ranked as easier in the SPM-C feature the relation in accordance with the reading direction: (from left to) Right or (from top to) Bottom. The former receives 0 and the latter 1 extra point for each row or column, respectively, in which the relation is correctly presented. The other directions occur in increasingly more difficult items of the SPM-C: the direction Left scores an extra 1, Top scores an extra 2, and Center an extra 2 per row or column.

*2.2.2.2. Components & Specifications.* The scores of Components & Specifications reflect divergent production. The Components score equals the number of different components or subcomponents in the completed matrix. To distinguish components from each other, we used the criterion that they should belong to distinct semantic or pictorial categories. This distinguishes Figurative components such as trees, flowers, etc., predominant in young children but also Non-figurative ones, mostly abstract geometrical shapes, such as crosses or circles. Depending on the degree of detail in a drawing, a tree could be scored for its subcomponents: e.g. branch, bird, separately. Crosses, circles, etc. may also occur as subcomponents of a component, in which case they are scored separately, using the same method as for components. To identify subcomponents, we used the criterion of separability (Garner & Felfoldy, 1970; Pomerantz, Sager, & Stoeber, 1977); cross and circle can exist as separate semantic or pictorial elements, whereas circle and its shading cannot. Subcomponents may, however, overlap in space, for instance a cross superimposed on a circle. We distinguished the following categories of components: Figurative; this category is, in principle, unbounded, and therefore we should not expect to arrive at a fixed set of subcategories; and Non-figurative, of which the items may be expected to fall into a limited number of different subcategories: Line, Square, Circle, Triangle, Diamond, Cross, Beam, Oval, and Dot.

*Specifications* scores features such as the shading used in drawing a component. The Specifications score equals to the number of different specifications used for depicting the (sub) components. Seven distinct categories of Specifications were applied. The first three categories are surface texture-specifications: *Black-White*, *Shaded*, and *Speckled*, and the fourth category: *Undulated*, applied to the way the outline of a component was rendered. Three further specifications may be considered as specifying a component relative to a *framework*: The fifth category: *Size*, applied to the rendering of any component or subcomponent in an unusually small

or large size. The sixth category: *Orientation*, applied when unrelated components or subcomponents varied in orientation across a row or column. The seventh category: *Number*, applied when a subcomponent was rendered more than once within a component. To demarcate Framework specifications from Relations, note that *Size*, *Orientation*, and *Number* applied only as long as they were ornamental and do not constitute a Relation.

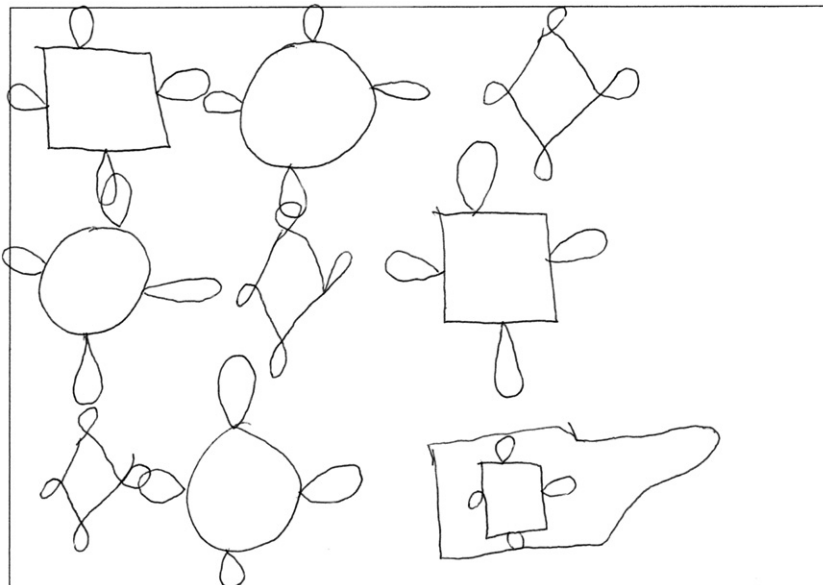
**2.2.2.3. Example.** We present a worked-out example of our scoring method at the hand of an item created by a Grade 1 girl, almost seven years old (Fig. 2). The item consists of squares, circles, and diamonds, each with four leaves. The completion in the lower-right corner is presented in a frame, drawn spontaneously by the child in recollection of preceding SPM-C items. The frame does not constitute a component or subcomponent and is therefore not taken into account in the scoring procedure. We start from the most encompassing relation, the one that applies to the largest subset of components, and confirm that it is covered by Succession. Accordingly, the top row of the item is represented as  $PQR$ . We assign the values  $P=1$ ,  $Q=2$ , and  $R=3$ . The next row reads  $QRP$ , in which  $Q$  scores 4 (2, the previous score for  $Q$ , +1 for new row, +1 for new column relative to the preceding row),  $R$  likewise scores 5 ( $3+1+1$ ) and  $P$  scores 3 ( $1+1+1$ ). Of the components in the last row, only the diamond matches the succession and is therefore scored with 7 ( $5+1+1$ ). The remaining components do not score extra points as they fall outside the scope of the relationship, presumably because an error was made. The score of the relation for this item equals 25 points, the sum of the values of all components. Had Succession expressed in the first two rows continued across the complete last row, the score for Succession covering the entire matrix would have been 36. Having scored the most encompassing relation, we seek whether further relations occur in

the item. The components on the diagonal contain diamonds, and therefore the item might qualify for a Group of Components. However, these items are not independent of the more encompassing Succession. Given that there are no further relations, the Relations score for this item remains 25.

Furthermore, all components have four leaves. As their number, size, and orientation are identical for all components, they do not contribute to the relations score. The same would have happened if the leaves had, e.g. been increased in size from  $P$  to  $Q$  and from  $Q$  to  $R$ . The extra variation is redundant, as both the relation type (Succession) and the operands ( $PQR$ ) remain the same. Had a different operation been used for the leaves (e.g. Change in size, or Indication of Mathematical Operation), it would have received an independent, additional score. Note that when exactly the same relation is applied two or more times (regardless of whether it is applied across rows or columns), no extra score is given.

*Components* scores four points, one each for separate components square, circle, diamond, and leaf. For *Specifications*, all except the leaves are drawn in simple, uniform lines and receive no extra point for texture, number, etc. The *leaves* receive an extra score of 1 on *Number* because leaves are applied more than once within components. The total score for *Components* & *Specifications* is therefore  $4+1=5$ . Let us assume that the child had used a bold line for the circle and a speckled line for the diamond. This would have resulted in an extra score of 2 on *Specifications*.

The scoring method reflects our theoretical assumptions about creative reasoning in relation to convergent and divergent operations: *Relations*, which values convergent production, gives higher scores to more, more complex, and more correct relations applied to the matrix. *Components* & *Specifications*, which scores for divergent production, yields higher scores for more, more detailed and more richly varied components. The fact that both are scored for the same



**Fig. 2.** Example of an item created in the CRT. Participant is an almost seven year old girl from Grade 1. The girl applied the relation Succession but not correctly over the complete matrix. With the new method we could scores the cells in which the relation was correctly applied.



creations has some notable consequences for the demarcation of convergent and divergent production. For instance, the leaves in Fig. 2 score only for Specifications, and thus qualify as divergent production; had they had any relation to each other, such as a systematic increase in number over columns (one, two, three leaves), they would have scored for Relations and had contributed to the convergent production score. This is in accordance with our view, that both divergent and convergent operations are part of the process of creative reasoning; that divergent precedes convergent operation in creative production, but that the latter uses only a subset of what is generated by the former in a creative synthesis (Jaarsveld & van Leeuwen, 2005). The drawback is that failed efforts to instantiate a Relation may unintentionally increase the Specifications score. We do not expect this to seriously bias our scores; the sub score Relations is picking up correct relations at the within-row or column level, and failed efforts at this level will be relatively rare.

### 2.2.3. TCT-DP

The TCT-DP was developed to assess an individual's "creative potentials", including personal and cognitive aspects. Urban and Jellen (1995) reported a test-retest reliabilities of  $r_s = .38$  after one and a half month and rank correlations of  $r_s = .38$  to  $r_s = .78$  ( $.05 > p > .001$ ) for different age ranges and time lapses. Urban (2005) observed that correlations with the Raven test showed values between  $r = .29$ ,  $p < .001$ , for children seven to ten years old ( $N = 190$ ) and  $r = .21$ ,  $p < .001$ , for participants from eleven to eighteen years old ( $N = 410$ ). Participants are presented with an unfinished drawing and are asked to complete it. The test form contains a large frame with five simple line figures within its border and a sixth figure on the outside. Participants are free to associate and elaborate on the given figures and to complete the drawing in any way they like. It is emphasized that there is no wrong way of completing the drawing. The administration of the test takes 15 min. For all participants in the age range of 5 to 95 years, the test can be performed individually or as a group. The test is supposed to be culture-fair. It embodies a holistic approach to creative production and emphasizes the shaping, production, and final "Gestalt" of the end product.

The TCT-DP score measures fluency and the qualitative attributes of a drawing, such as organization and elaboration (Urban, 2005). In other words, it measures divergent thinking as we understand it. The score is based on the elaboration of associations, and the number, originality, and organization of ideas. The TCT-DP does not, however, contain a sub score specifically defined for measuring elaboration. The score is composed of 14 sub scores: *Continuation*, *Completion*, *Connections Made with a Line*, *Connections Made to Produce a Theme*, *Figure-based Boundary Transgression*, *Figure-Independent Boundary Transgression*, *Perspective*, *Humor and Affectivity*, *Unconventionality-a*: any manipulation of the material; *Unconventionality-b*: any surrealist, fictitious and/or abstract elements or drawings; *Unconventionality-c*: any usage of symbols or signs; *Unconventionality-d*: unconventional use of given fragments; *Speed*: drawings that are made within a certain time limit and show a score above a certain value score extra points. All sub scores have a maximum of six points.

The manual gives no validity for the separate sub scores but gives the results of a factor analysis; the sub scores *New Elements*, *Connections Made with a Line*, *Connections Made to Produce a Theme* and *Perspective* loaded significantly on Factor 1 The Drawing of a Total Composition; the sub scores *Continuation*, *Completion* and *Figure-based Boundary Transgression* loaded significantly on Factor 2 The Use and Assimilation of given Components and the sub scores *Humor* and *Unconventionality-a,b,c,d* loaded on Factor 3 Unconventionality.

### 2.3. Procedure

We administered the tests to groups of seven to nine children in a room provided by the school. Children were sitting at least 1 m apart from another, minimizing disturbance. Two test assistants supervised the sessions and saw to it that the children completed each form. They explained the complete procedure and what should be expected; they handed out the forms and introduced each test according to the manual. Ballpoint pens were used to ensure that no information was lost through erasing. When a child had completed a test form, home work was done until the time allowed for the test had elapsed. For the SPM-C this time was approximately 45 min, for the CRT 20 min, and for the TCT-DP 15 min. Testing was performed only in-between the regular school breaks. The total test session comprised about 100 min, breaks included.

## 3. Results and discussion

In contrast to the SPM-C and TSD-DP, the CRT does not offer reliability criteria yet. To obtain an indication of the reliability of our scoring method, a subset of 69 items was scored independently by different raters. This resulted in an inter rater reliability of  $r = .99$ ,  $p < .01$  for the CRT, and of  $r = .90$ ,  $p < .01$  (both 2-tailed), for the TSD-DP. When considering CRT sub scores, we found an inter rater reliability of  $r = .99$ ,  $p < .01$  for the CRT sub score Relations, which measures convergent thinking, and a reliability of  $r = .91$ ,  $p < .01$  (both 2-tailed) for the CRT sub score of Components & Specifications, which measures divergent production. Note that the reliability for the divergent score of the CRT is similar to that of the TSD-DP. Interrater reliability over the type of relations was calculated with Cohen's *Kappa*,  $K = .936$ . Table 2 gives the frequencies of application ( $n$ ) of the relations. Of all participants ( $N = 205$ ), nine (4.4%) produced a matrix that could not be scored with any relation which resulted in a score of zero; 179 (87.3%) applied one relation, eleven (5.4%) applied two relations, and six (2.9%) applied three relations; Participants of Grade 1 applied one relation (94%), those of Grades 2, 3, and 4 applied up to three relations with a maximum for three relations in Grade 4 (5.0%).

### 3.1. Scores by grade level

The mean scores of SPM-C, TCT-DP, CRT Relations and CRT Components & Specification, as well as the CRT total score (Relations + Components & Specification) are displayed in Table 3, separately for each grade level. Effects of grade level were found to be significant for SPM-C  $F(3,201) = 32.400$ ,  $p < .01$ ,

and TCT–DP  $F(3,201) = 4.350, p < .01$ , but not for CRT total and neither for the two sub scores of the CRT. Bonferroni post-hoc analysis revealed a significant increase in the SPM–C from Grade 1 to Grades 2–4 ( $p < .001$ , respectively), and from Grade 2 to Grade 4 ( $p < .001$ ). There is an increase tendency at the 10% level from Grade 2 to Grade 3 ( $p = .097$ ). No increase was found from Grade 3 to Grade 4 ( $p > .1$ ). For the TSD–DP the only significant increase was from Grades 1 to 3 ( $p < .05$ ). For the CRT, neither increases nor decreases in scores were found across grades, not in its total score, nor for any of its subscores.

3.2. Correlations

As common in the literature, we calculated Pearsons' correlation coefficients (e.g., Barron & Harrington, 1981; Guthke, 1992; Preckel et al., 2006). They have greater statistical power than nonparametric correlations (Spearman Rho). However, existing models from test theory (e.g., Guthke, 1999), which would validate the use of parametric statistics are not readily applicable here. We therefore report significance levels using two  $p$  values:  $p_n$  representing the significance level under the classical assumption of the interval level samples approaching the normal distribution, and  $p_p$  based on permutation methods (Efron, 1979, 1981), which are simple and free of assumptions about the distribution of the statistic.

Our correlations are reported in a hierarchical fashion: firstly, a limited number of correlations relevant to the main issues of the paper: these are the correlations amongst the total scores of three tests and of these with the subtests of the CRT, the CRT Relations and CRT Components & Specifications assumed to measure, respectively, convergent and divergent operations. For exploratory purposes only, we followed this up with correlations between sub scores of the TCT–DP and specific relations.

The sub score of CRT Relations is based on the application of any of the fifteen defined relations. Almost all children ( $n = 179, 87.3\%$ ) received a score for the application of at least one relation (see Table 4). Correlations included the children ( $n = 9$ ) who made an item which, however, could not be scored. Significance levels were corrected for multiple comparisons using Bonferroni correction.

Main correlations are shown in Table 5. The CRT total score correlated with the SPM–C score ( $r = .183, p_n < .01, p_p < .01$ ) but

**Table 4**  
Number of applications (percentages) of relations in the CRT.

Grade	Application			
	0	1	2	3
all ( $N = 205$ )	9 (4.38)	179 (87.32)	11 (5.37)	6 (2.93)
1.	3 (5.88)	48 (94.12)	0	0
2.	3 (6.98)	36 (83.71)	3 (6.98)	1 (2.33)
3.	2 (3.92)	41 (80.39)	6 (11.77)	2 (3.92)
4.	1 (1.67)	54 (90.00)	2 (3.33)	3 (5.00)

Note. Most children ( $n = 179$ ) applied one relation in their matrix. The most frequent application of two relations ( $n = 6$ ) is observed in Grade 3.

not with the TCT–DP score, which, in turn showed a correlation with the SPM–C score ( $r = .225, p_n < .001, p_p < .001$ ). Partial correlation analysis revealed correlations of CRT and SPM–C corrected for TCT–DP ( $r = .197, p < .05$ ), and of SPM–C and TCT corrected for CRT ( $r = .237, p < .001$ ). No correlation was found between CRT and TCT–DP corrected for SPM–C.

As expected, the CRT sub score Relations, which according to our theory represents convergent thinking, showed a correlation with the SPM–C ( $r = .192, p_n < .01, p_p < .01$ ; partial correlation corrected for TCT–DP:  $r = .213, p < .01$ ) and the CRT sub score Components & Specifications, which according to our theory represents divergent thinking, showed a correlation with the TCT–DP ( $r = .147, p_n < .05, p_p < .05$ , partial correlation corrected for SPM–C:  $.153, p < .05$ ). Furthermore, as expected from the assumption that both sub scores of the CRT represent different thinking abilities, no correlation between them was found. Moreover, as expected, there were no correlations between CRT–Relations and TCT–DP, and of CRT–Components & Specifications and SPM–C. The latter results hold also for partial correlation analyses.

The threshold theory assumes that correlations between creativity scores and IQ are limited to individuals with an  $IQ \leq 120$ . In our sample, there are only 8 individuals with an  $IQ > 120$ . For this reason, our data did not permit a test of the threshold theory.

The correlations observed offer evidence towards convergent validation of the CRT test scores. This, even though they are only somewhat lower than those obtained elsewhere (e.g. Preckel et al., 2006, has correlations between IQ and figural creativity of .09–.25 and with numerical creativity of .20–.27). High correlations are not what we would have the-

**Table 3**  
Scores per grade of psychometric measures.

Grade	Means (SD)				
	SPM–C <sup>a</sup>	TSD–D <sup>b</sup>	CRT <sup>c</sup>	CRT–R <sup>d</sup>	CRT–CS <sup>e</sup>
1	27.04 (12.74)	12.30 (5.48)	12.88 (09.61)	7.21 (09.63)	5.67 (2.47)
2	33.78 (8.97)	12.46 (5.12)	14.98 (12.51)	9.72 (11.81)	5.26 (2.85)
3	38.91 (7.10)	14.41 (5.30)	16.88 (14.16)	10.73 (13.79)	6.15 (3.04)
4	41.39 (7.53)	14.68 (4.64)	13.28 (10.17)	7.88 (09.92)	5.40 (2.40)

Note. The mean scores of the SPM–C differed between Grades 1 and 2; those of the TCT–DP differed between Grades 1 and 3. No differences were observed for the mean scores of the CRT (Post hoc test, one way ANOVAs, Bonferroni, alpha .05).

- <sup>a</sup> Standard Progressive Matrices–Classic.
- <sup>b</sup> Test for Creative Thinking–Drawing Production.
- <sup>c</sup> Creative Reasoning Task (sum of sub scores Relations and Components & Specifications).
- <sup>d</sup> CRT Relations, sub score on convergent thinking.
- <sup>e</sup> CRT Components & Specifications, sub score on divergent thinking.

Please cite this article as: Jaarsveld, S., et al., Creative reasoning across developmental levels: Convergence and divergence in problem creation, *Intelligence* (2012), doi:10.1016/j.intell.2012.01.002

**Table 5**  
Main correlations of psychometric measures.

Measure	1	2	3	4	5
1. SPM-C	–	.225***	.183**	.192**	–.010
2. TCT-DP		–	–.035	–.070	.147*
3. CRT			–	.973**	.245**
4. CRT-R				–	.016
5. CRT-CS					–

Note. According to our hypothesis the scores of the CRT Relations correlated with those of the SPM-C, CRT Components & Specifications correlated with TCT-DP, while both sub scores showed no relation. Pearson correlations, 2-tailed.

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

oretically predicted. We have observed elsewhere (Jaarsveld et al., 2010) that creative reasoning differs from classical reasoning. With our current method it is impossible to calculate discriminant validity in terms of inter-item correlations, since there is only one item in the CRT. We observe, however, that the sub scores of the CRT that measure convergent and divergent operations in creative reasoning, respectively, are highly selective in their correlation with, respectively, SPM-C and TCT-DP. In addition, they indirectly offer some evidence towards comparing the discriminant validity of our measures to those of SPM-C and TCT-DP. These two tests show the highest partial correlation, even though they are supposed to measure different constructs.

### 3.2.1. CRT sub-scores and TCT sub-scores

For exploratory purpose we correlated the sub scores of the CRT and the sub scores of the TCT-DP with each other and with the scores of the SPM-C. The relation Combination showed a correlation with the TCT-DP sub scores Figure-based Boundary Transgression and the relation Groups of Components showed a correlation with TCT-DP sub scores Unconventionality-b ( $r = .281, p_n < .01, p_p < .01$ , and  $r = .458, p_n < .01, p_p < .01$ , respectively). Both results might indicate that an unconventional approach is relevant to the generation of structure in a matrix. Combination was the only relation that showed a correlation with the SPM-C score ( $r = .164, p_n < .01, p_p < .01$ ). This result might be due to the observation that Combination was the only relation showing a significant increase over grade levels (see paragraph Relations featured, solved and applied, Quantitative). The TCT-DP sub scores Completion and Connections Made with a Line correlated with the SPM-C score ( $r = .232, p_n < .01, p_p < .01$ , and  $r = .163, p_n < .01, p_p < .01$ , respectively). The CRT sub score Components & Specifications correlated with the TCT-DP sub scores Completion and Unconventionality-d ( $r = .222, p_n < .01, p_p < .001$ ; and  $r = .144, p_n < .05, p_p < .002$ , respectively).

### 3.2.2. Gender differences

We observed a gender effect in the CRT,  $t = 2.248, df = 203, p < .05$ : total mean score was higher for girls ( $M = 16.28, SD = 12.212$ ) than for boys ( $M = 12.64, SD = 10.905$ ). Sub scores relating to convergent and divergent thinking individually were only marginally significantly favoring girls. Texture Completion was the only specific CRT Relation for which a difference between girls ( $M = .24, SD = .428$ ) and

boys ( $M = .42, SD = .496$ ) were observed: ( $t = 2.862, df = 203, p < .01$ ). Note, however, that the large SD suggests outliers, so this result may be spurious. We observed no differences between girls and boys in the SPM-C and a marginal effect of gender in the TCT-DP ( $t = 1.925, df = 203, p = .056$ ): higher scores for girls ( $M = 14.44, SD = 5.49$ ) than for boys ( $M = 13.06, SD = 4.74$ ). The TCT-DP sub score humor ( $t = 2.803, df = 203, p < .01$ ) favored girls ( $M = 0.58, SD = 0.89$ ) over boys ( $M = .28, SD = .66$ ). Again, these results may be spurious.

**3.2.2.1. SPM-C and CRT.** Girls showed a correlation of the SPM-C score with the CRT total ( $r = .234, p_n < .05, p_p < .05$ ), and with its subscore Relations ( $r = .254, p_n < .01, p_p < .05$ ). We observed a positive correlation for the more complex relation Group of Components ( $r = .197, p_n < .05, p_p > .05$ ), and negative correlations for Idiosyncratic and Semantic Coherence ( $r = -.355, p_n < .01, p_p < .01$ ), and Iteration of Two Components ( $r = -.229, p_n < .05, p_p < .01$ ). All measures in the CRT that correlate with the SPM-C in Girls are associated with convergent thinking. The measure Components & Specifications, which is the one associated with divergent thinking, did not correlate with the SPM-C in girls ( $r = -.065$ ). For Boys the correlation between the SPM-C score and the CRT total was not significant ( $r = .134$ ). The CRT sub scores Relations did not show any significance with the SPM-C scores either ( $r = .126$ ). Amongst the specific relations, Pattern Completion showed a positive correlation ( $r = .237, p_n < .05, p_p < .01$ ) and negative correlations were found with Iteration of One Component and Change & Increase ( $r = -.322, p_n < .01, p_p < .01$  and  $r = -.224, p_n < .05, p_p < .01$ , respectively). The sub score Components & Specifications did not yield significance ( $r = .052$ ). Female participants who were able to infer relations between components in well defined problems were also good in applying relations in ill defined problems. This implies that in female participants there is overlap between classical reasoning and the convergent component of creative reasoning. For male participants, these reasoning abilities seem to differ more radically. The creative reasoning sub scores associated with divergent reasoning are correlated neither in Boys nor Girls with classical reasoning scores.

**3.2.2.2. CRT and TCT-DP.** In Girls there was no correlation of the overall scores of the TCT-DP with the CRT, nor did we with the CRT sub score Relations associated with convergent thinking. The sub score Relations correlated with the TCT-DP sub score Unconventionality-d ( $r = .254, p_n < .01, p_p < .01$ ). The CRT sub score associated with divergent operations, Components & Specifications, did not correlate with the TCT-DP score overall, but correlated with the TCT-DP sub scores Unconventionality-d ( $r = .254, p_n < .01, p_p < .01$ ) and Continuation ( $r = .255, p_n < .01, p_p < .05$ ). The significant correlations in Girls of TCT-DP sub score Unconventionality-d with both sub scores of the CRT underline the creative character of the latter task especially for the convergent thinking ability.

In Boys the CRT showed no relation of with the TCT-DP overall; neither did the CRT sub score Relations associated with convergent operations. There was, however, a negative correlation of CRT Relations with the TCT-DP sub score

Connections Made to Produce a Theme ( $r = -.201, p_n < .05, p_p < .01$ ). Boys showed a relation between the CRT sub score associated with divergent production, Components & Specifications, and the TCT–DP score overall ( $r = .210, p_n < .05, p_p < .01$ ) and with its sub scores Unconventionality-d and Humor, (respectively,  $r = .278, p_n < .01, p_p < .01$  and  $r = .208, p_n < .05, p_p < .05$ ). A significant correlation of the TCT–DP overall with the CRT sub score of Components & Specifications was present in the complete sample but the results show that it is restricted to Boys. This correlation underlines the divergent thinking aspect of the creative thinking task.

**3.2.2.3. SPM-C and TCT–DP.** Girls showed a correlation of the SPM-C with the TCT–DP overall ( $r = .289, p_n < .01, p_p < .01$ ) as well as with the sub scores Completion ( $r = .229, p_n < .05, p_p < .05$ ). Boys showed no correlation of the SPM-C with the TCT–DP overall ( $r = .158$ ) but with its subscore Completion ( $r = .239, p_n < .05, p_p < .05$ ). This result could be explained by the fact that completion also is performed in the SPM-C. In other words, the correlation may indicate a common characteristic of completion tasks regardless of intelligence or creativity. The correlation between SPM-C and TCT–DP found in girls was also observed in the complete sample. The present results show that they are restricted to girls.

We further observed an unexpected correlation between TCT–DP and SPM-C scores. This result appeared to be restricted to girls and their TCT–DP sub score Completion. No difference between male and female participants was observed for the other measures.

### 3.2.3. Figurative versus Non-figurative

We observed little overlap between children using Figurative and non-Figurative components. Only nine children used both types of components, they were excluded from the following analyses. We categorized the generated items of the CRT according to the Non-figurative ( $n = 140$ ) versus Figurative ( $n = 56$ ) character of the representation applied (see Components & Specifications in the Material section). We observed a difference in mean scores of the SPM-C ( $t = 2.708, df = 194, p < .01$ ) between participants who used Figurative and Non-figurative components: a higher score for participants who used Non-figurative components ( $M = 38.11, SD = 9.172$ ) than for those who used Figurative components ( $M = 33.95, SD = 10.970$ ). However, there was no difference between the use of Non-figurative ( $M = 13.20, SD = 4.789$ ) and Figurative ( $M = 14.80, SD = 5.626$ ) components in the mean scores of the TCT–DP ( $t = 2.012, df = 194, p = .064$ ). From these results we may infer that the ability to generate relations between Non-figurative components, as contrasted to Figurative components, is related to the ability to reduce relations from between Non-figurative components (SPM-C) and is not related to the ability to enhance and complete simple line components (TCT–DP).

The CRT sub score Relations, scoring for convergent abilities, was higher on average for participants who used Non-figurative components ( $M = 10.28, SD = 11.901$ ) than for those who used Figurative components ( $M = 4.80, SD = 8.740; t = 3.121, df = 194, p < .01$ ). By contrast, the CRT sub score Components & Specifications, scoring divergent abilities operating in the CRT, was higher on average for participants who used Figurative components ( $M = 6.96,$

$SD = 3.133$ ) than for those who used Non-figurative components ( $M = 4.99, SD = 2.292; t = 4.891, df = 194, p < .01$ ). Participants who did well in reducing relations amongst Non-figurative components, therefore, also did well in reasoning creatively with those types of components.

We might infer that participants, who tend to apply Non-figurative components in an ill defined problem situation, have a better developed ability to reason about relations between components. By contrast, participants who tend to apply Figurative components would have a more pronounced ability to pay attention to form, detail and design. Note that these results are valid, at best, only within the context of the theory on which the present scoring method is based; as observed in the Methods section, inevitably in some cases failed attempts to produce a Relation will positively contribute to the Components & Specifications score. There was a difference in age between participants who used Figurative and Non-figurative components ( $t = 2.139, df = 194, p < .05$ ). Participants who used Non-figurative components were older on average ( $M = 8.76$  y,  $SD = 13.416$ ) than those who used Figurative components ( $M = 8.35$  y,  $SD = 13.416$ ). From this we may infer that within our sample the use of Non-figurative components increases with age.

### 3.3. Relations featured, solved and applied

Jaarsveld et al. (2010) introduced a set of predefined relations for Raven matrices, and compared the frequency distributions of the relations featured and solved in the SPM-C, with those applied in CRT creations. They found the frequency distributions in the CRT to differ from those in the SPM-C and concluded that different forms of reasoning were measured by both tasks. Here we replicate the analysis from Jaarsveld et al. (2010) based on our current data and our present, more differentiated categorization and score system for the relationships.

For our present relation categories, we analyzed which relation featured in each SPM-C item and which occurred in the items a child with a given norm score is expected to have solved, as well as which rules were applied in the items created in the CRT (Table 6). Noticeable is, firstly the high frequency in Applied as opposed to Solved, of Relation 1, Idiosyncratic Coherence, as well as both relations of the String format, and of Relation 3, Pattern Completion in Grades 2–4. Noticeable, secondly, are the low frequencies in Applied of Relation 6, Symmetry and Relation 7, Change & Increase.

These results might seem to imply a dependency of the CRT on the SPM-C, in that Rules Applied appear to be less complex than the ones Featured and Solved. Indeed, the simpler rules are, on average, more prominent in the former. On the other hand, older children tend to use rules of similar complexity in SPM-C and CRT. The simplest rule: Idiosyncratic Coherence shows a prominence only in the youngest children. Moreover, the other simple rule, Texture Completion, was used in the SPM-C in Non-figurative items, but featured in the CRT mostly in Figurative designs. With increasing grade level this relation showed a decrease of the frequencies for Solved in the SPM-C but an increase in the frequency of Applied in the CRT (see Table 6). Older children tend to use rules of similar complexity in SPM and CRT. In Jaarsveld et al. (2010), we searched exhaustively for dependencies

**Table 6**  
Frequencies in percentages per grade for relations featured and solved in the SPM-C and applied in the CRT.

Grade	Relations														
	Continuous Pattern					String					3 × 3 Matrices				
	Idiosyncratic Coherence	Jigsaw form	Pattern Completion	Iteration of one Component	Iteration of two or >2 Components	Symmetry	Change & Increase	Succession 2 × 2	Change	Increase	Succession 3 × 3	Combination	2 Values	Contrast	Groups of Components
1	0.00	0.00	10.53	0.00	0.00	7.02	18.42	0.00	14.04	15.79	12.28	9.65	7.02	3.51	1.75
2	0.00	0.00	22.24	0.00	0.00	12.65	16.85	0.00	11.92	20.14	7.49	8.17	0.46	0.00	0.09
3	0.00	0.00	17.91	0.00	0.00	12.89	18.06	0.00	13.44	20.71	8.62	8.14	0.24	0.00	0.00
4	0.00	0.00	16.18	0.00	0.00	11.48	20.42	0.00	12.77	19.29	9.72	9.52	0.58	0.00	0.06
Applied	0.00	0.00	15.31	0.00	0.00	10.77	20.42	0.00	13.21	18.23	10.08	9.40	2.01	0.23	0.36
1	33.33	3.92	5.88	9.80	11.76	1.96	1.96	1.96	5.88	3.92	13.73	1.96	0.00	0.00	3.92
2	19.15	2.13	25.53	0.00	4.26	0.00	4.26	2.13	10.64	10.64	8.51	6.38	0.00	0.00	6.38
3	17.24	0.00	32.76	5.17	6.90	1.72	3.45	0.00	6.90	5.17	3.45	5.17	0.00	0.00	12.07
4	11.48	0.00	55.74	0.00	1.64	1.64	0.00	3.28	1.64	1.64	6.56	8.20	0.00	0.00	8.20

Note. Between Featured, Solved and Applied, Chi square values for independence were significant for all grades between Solved and Applied, and for none of the grades between Featured and Solved (see Table 7).

between SPM-C and CRT in terms of developmental lag, but found no systematic dependencies.

3.3.1. Qualitative

We considered whether Relations in the items created in the CRT reflected those featured and solved in the SPM-C. According to Table 6, we observed that the children in Grade 1 applied Relation 1, Idiosyncratic Coherence, with the highest frequency. Relation 1 is still frequently applied in Grades 2 to 4, but now the highest frequencies are observed for Relation 3, Pattern Completion. An identical shift in relation preference was observed in Jaarsveld et al. (2010). Four of the five relations that featured in the CRT but not in the SPM-C showed a decreasing trend over grade levels (Relation 1, Idiosyncratic Coherence; Relation 2, Jig saw; and Relations 4 and 5 from the string format), and one showed an increasing trend (Relation 8, Succession 2×2).

3.3.2. Quantitative

For each grade level, Table 7 provides percentages of the total number of Relations featured in the SPM-C, solved in the SPM-C and applied in the CRT. For the Chi Square calculations, we only included those Relations that showed a relative frequency of more than five percent for featured, solved, and applied. Note that with this criterion, all relations that are exclusive to either SPM-C or CRT automatically drop out from the analyses, meaning that subsequent tests are biased in favor of sameness of frequencies between SPM-C and CRT. The relations in the formats Continuous Pattern and String are taken together because separately they featured either in featured and solved or in applied. Despite the bias in the test, the distribution of relations Applied differed from Solved for all grades, with min  $\chi^2(2) = 17.23, p < .001$  (calculated over Relations 1–3, 11 and 12) for Grade 4, and max  $\chi^2(4) = 31.40, p < .001$  (calculated over Relations 1–3, 9, 10, 11 and 12) for Grade 3; Applied also differed from Featured; for all grades, here we observed a min  $\chi^2(4) = 23.30, p < .001$  (calculated over Relations 1–3, 9, 10, 11 and 12) for Grade 2, and a max  $\chi^2(4) = 39.49, p < .001$  (calculated over Relations 1–3, 9, 10, 11 and 12) for Grade 3; Featured and Solved in the SPM-C did not differ in any of the school grades.

With increasing grade level, the total frequencies per grade of two relations in the solving condition showed a decrease, i.e., Relation 6, Symmetry and 10, Increase,  $r_s = -1.00, n = 4, p < .01$ , for both relations. In the same condition three relations showed an increase, i.e., Relations 7, Change & Increase, 9, Change and 11, Succession 3×3,  $r_s = 1.00, n = 4, p < .01$ , for all three relations. Note that Relation 3, Pattern Completion, showed a decrease in the solving condition and an increase in the applied condition  $r_s = -1.00, n = 4, p < .01; r_s = 1.00, n = 4, p < .01$ , respectively, whereas Relation 12, Combination, showed an increase in both conditions, solving:  $r_s = 1.00, n = 4, p < .01$ , and application:  $r_s = .949, n = 4, p < .05$ .

Altogether, these observations corroborate our hypothesis that reasoning abilities lead to different performances when applied in solving well defined problems (classical reasoning) then when applied in solving ill-defined problems (creative reasoning).

#### 4. General discussion

We used the Creative Reasoning Test (CRT) to measure processes of convergent and divergent reasoning, which we consider to play a role in open problem solving processes, and compared CRT scores with classical reasoning and creativity tests.

##### 4.1. Grade level

The CRT scores, remarkably, did not increase with grade levels. One could argue that our participants did not have yet enough expertise to show a developmental course. However, the level that we were assessing is what *Boden (1990)* and *Barsalou and Prinz (1997)* defined as P-creativity and Mundane creativity, respectively. Research into creative cognitive processes related to non expert individuals is in line with the Creative Cognition Approach (*Smith, Ward, & Finke, 1995*). The finding of an absent of increase in CRT scores is consistent with earlier observations which failed to find a relation between creativity and age (*Ward, 1968*). However, development does not stop at the end of primary school. For this reason we are investigating the development of creative reasoning in secondary schools.

The absence of an increase in CRT scores contrasts with that of the classical reasoning test (SPM-C), where we observed continuous increase in scores with grade. There are several explanations for this discrepancy. One is that the SPM-C is based more on abilities that are trained in school, for instance logical reasoning is trained in mathematics. The creative reasoning abilities tested in the CRT are perhaps less systematically trained in school. A second possible explanation resides in the construction of the SPM-C. In the SPM-C, items increase in difficulty from one to the next, and each new item builds upon apprehension of the previous one. Because of training, increased memory, attentional and other skills, older children may accumulate this information more effectively than younger ones. A third explanation is that intelligence is subject to development to a greater extent than the creative abilities currently under study. This squares

with the observation of creativity in the early stages of development (*Cramer, 1995*) and its remarkable resilience in old age (*Simonton, 1990*). Finally we must consider the current procedure, in which after several SPM-C items, only a single CRT item was required. This was done in order to tap individual abilities at the moment where they had reached the maximum level of apprehension according to SPM-C. However, children may not achieve to their maximum abilities in this single item. For this reason, we are currently investigating the effect of including multiple CRT items in the task.

##### 4.2. Correlations

Of the two sub scores of the CRT, one: Relations, a measure of convergent production in ill-defined problem spaces, correlated with the score on the SPM-C, a classical reasoning test. The other CRT sub score: Components & Specifications, a measure of divergent production, correlated with the score of the TCT-DP, a creativity test. The present study, therefore, corroborates our hypothesis that creative reasoning involves both convergent and divergent production, and that these are what our respective sub scores measured. Other observations that further strengthen our hypothesis are the contrasting correlations of TCT-DP with SPM-C and CRT. In our study as well as in earlier studies a correlation between TCT-DP and SPM-C was observed (*Urban, 2005; Urban & Jellen, 1995*), indicating a relation between a classical creativity and intelligence test. Importantly, however, the divergent and convergent sub scores of the CRT did not show a relation.

The CRT measures convergent as well as divergent abilities in one and the same creative reasoning problem. This allowed us to meaningfully compare the individual sub scores for convergent and divergent abilities. We found no correlation whatsoever between the sub scores. From this we may conclude that, even though creative reasoning may require interplay of divergent and convergent operations, they are clearly distinguishable abilities in this task. Some issues remain open to debate: for instance, when participants employ multiple relations this increases the convergent score, but this could also be viewed as divergent performance. In our current theoretical framework, however, convergent production includes control operations required to integrate multiple ideas (see also *Jaarsveld & van Leeuwen, 2005*).

The observed relationship between the CRT subscore Relations and the score of the SPM-C highlights the role of convergent thinking abilities in ill defined problem situations. Convergent abilities involved in both CRT and SPM-C operate on an identical knowledge domain; both tests instructed participants to work with the same kind of components and their relations. They differ, however, in their problem spaces. As a result, both tasks show contrasting preferences for certain Relations across all school levels.

The solving task serves, amongst other things, to make participants acquainted with the particular structure of the matrices problems. Without this phase we would have needed extensive instructions, which might decrease creative production. *Amabile (1987)* gave one group specific instructions on how to produce a collage. She points out that this renders the task algorithmic, and that it is therefore, inappropriate to

**Table 7**

Chi square values for the differences in frequency distributions of relations between featured, solved, and applied.

Condition	Chi <sup>2</sup>	df	<i>a</i>	Relation in calculation*
Grade 1				
Featured–Solved	7.14	6	n.s.	1–3, 6, 7, 9, 10, 11, 12
Featured–Applied	29.13	3	<.001	1–3, 9, 10, 11
Solved–Applied	21.34	3	<.001	1–3, 9, 10, 11
Grade 2				
Featured–Solved	4.34	6	n.s.	1–3, 6, 7, 9, 10, 11, 12
Featured–Applied	23.30	4	<.001	1–3, 9, 10, 11, 12
Solved–Applied	25.23	5	<.001	1–3, 7, 9, 10, 11, 12
Grade 3				
Featured–Solved	2.37	6	n.s.	1–3, 6, 7, 9, 10, 11, 12
Featured–Applied	39.49	4	<.001	1–3, 9, 10, 11, 12
Solved–Applied	31.40	4	<.001	1–3, 9, 10, 11, 12
Grade 4				
Featured–Solved	2.37	6	n.s.	1–3, 6, 7, 9, 10, 11, 12
Featured–Applied	26.78	2	<.001	1–3, 11, 12
Solved–Applied	17.23	2	<.001	1–3, 11, 12

Note. \* Only those relations were part of the Chi-squares that showed a frequency of more than five percentage points for Featured, Solved, and Applied.

assign the label creative to the performance of this specific group. To have a solving task precede a generation task is consistent with the general observation that nothing is invented from scratch; creativity implies using old elements in new contexts and seeing relations that no one saw before (Barron, 1981; Boden, 1990; Indurkha, 1992; Torrance, 1987). Familiarity with the relevant domain and experience with a variety of methods are a prerequisite for generating solutions (Voss & Post, 1988).

The relation between the CRT subscore Components & Specifications and the score of the standard creativity test, the TCT–DP, emphasizes the role of divergent production in creative reasoning. Although both tests present participants with ill defined problems, the knowledge domains they represent are different; in the CRT participants are asked to generate figures and their relations, while in the TCT–DP, participants are asked to complete a drawing with given figures by association. However, whether a figure is given or generated, the ability to complete and connect is relevant in both situations. We further observed an unexpected correlation between TCT–DP and SPM–C scores. This result appeared to be restricted to girls.

#### 4.3. Gender

A number of studies on differences in creativity as a function of gender in the 60s showed no significant differences by sex (Cronbach, 1968; Ward, 1968). Later studies showed differences in creativity between girls and boys, depending on age and on what measures are considered (e.g., Lau & Cheung, 2010). Other studies emphasized the importance of social factors such as social role playing. At five years old, girls who engaged more in early role play than boys also were found to score higher on the TCT–DP during early adolescence (Mullineaux & Dilalla, 2009). Boys with mixed sex-role patterns (inconsistent orientation and preference) scored significantly higher on creativity measures such as fluency and uniqueness than boys with consistent sex-role patterns (Biller, Singer, & Fullerton, 1969). Such gender differences may explain the higher test scores for girls than boys observed in our study. Interesting enough, however, independent from diagnostic scoring, psychophysiological studies suggest different hemispheric organization in men and women during divergent creative thinking (Razumnikova, 2004).

The CRT scores show same tendencies for the sub scores relating to convergent and divergent operations. More interesting, perhaps, are the contrasting results we observed between boys and girls in their respective sub scores for convergent and divergent processes. In girls the convergent component of the creative reasoning ability showed a relationship with classical reasoning ability. Such a relation was not observed in boys. On the other hand, the divergent component of the creative reasoning ability showed a stronger relationship with creativity tests in boys than in girls. This implies that the convergent operations used by girls are more domain-general than those of boys, whereas in boys it is divergent production which is more domain-general. The results, therefore point to an interesting gender difference in creative reasoning style: girls appear to rely more on abstractions in the reasoning aspects of the task,

and boys more on general principles in the production stages of creative reasoning.

#### 4.4. Figurative

A comparison between individuals with Figurative versus Non figurative preferences showed that the latter was related to higher scores on the SPM–C, but also that the underlying factor is school grade. Participants from the higher grades showed higher scores on the SPM–C, together with an increased application of non figurative components in the CRT.

#### 4.5. Relations

The present study used an enhanced version of the score method applied in Jaarsveld et al. (2010). The new method refined the consideration given to relations occurring in the matrices created, by taking their complexity into account. In addition, the method independently values the components to which these relations are applied. This resulted in two independent measures, which reflect, respectively, convergent and divergent aspects of creative reasoning. Scores on these two measures were uncorrelated.

#### 4.6. Limitations and conclusions

The CRT is in an early stage of development. In its current form, there are several issues that restrict its practical utility. For instance that before the CRT can be administered the SPM–C has to be completed. For the current study, SPM–C data were needed anyway. If one is interested only in CRT performance, however, future developments of the test should include a certain number of Raven matrices, specifically constructed to reflect the specter of relevant relations in the SPM–C. Moreover, the current CRT asks participants to generate one item only. Subsequent versions should consist of a specified, larger number of items. Despite these restrictions, the current results show that the capacities to solve ill defined problems can be contrasted to those that can be inferred from standard intelligence tests. Results showed that in each grade, the relations preferably applied in the Creative Reasoning task contrasted with those featured in the Standard Progressive Matrices test. Thus we may conclude that reasoning is broader than what is measured in a classical intelligence test, reasoning abilities lead to different performances when applied in solving well defined problems then when applied in solving ill defined problems.

#### Acknowledgments

We would like to thank Marisete Welter (University of Kaiserslautern, Germany) for providing the scores for the inter rater reliability calculations, and Frits Peters who wrote the software to compute the frequency profiles and the software for the permutation method. Many thanks are due to four anonymous reviewers for very helpful comments and remarks to an earlier draft of this paper. Many thanks are also due to the director, teachers and pupils of the Pestalozzi Gundschule Kaiserslautern for supporting this study.

## References

- Akin, O., & Akin, C. (1998). On the process of creativity in puzzles, inventions, and designs. *Automation in Construction*, 7, 123–138.
- Amabile, T. M. (1987). The motivation to be creative. In S. G. Isaksen (Ed.), *Frontiers of creativity research: Beyond the basics* (pp. 223–254). Buffalo, NY: Bearly Limited.
- Amabile, T. M. (1990). Within you, without you: The social psychology of creativity, and beyond. In M. A. Runco, & R. S. Albert (Eds.), *Theories of creativity* (pp. 61–91). Newbury Park, CA: Sage.
- Barron, F. (1981). Originality in relation to personality and intellect. In W. E. Barbe, & J. S. Renzulli (Eds.), *Psychology and education of the gifted* (pp. 168–178). New York, NY: Irvington Publications.
- Barron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology*, 32, 439–476.
- Barsalou, L. W., & Prinz, J. J. (1997). Mundane creativity in perceptual symbol systems. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 267–307). Washington, DC: American Psychological Association.
- Basadur, M. (2005). Modelling applied creativity as a cognitive process: Theoretical foundations. *Korean Journal of Thinking & Problem Solving*, 15, 13–41.
- Bear, J. (2003). Evaluative thinking, creativity, and task specificity: Separating wheat from chaff is not the same as finding needles in haystacks. In M. A. Runco (Ed.), *Critical creative processes* (pp. 129–151). Cresskill, NJ: Hampton Press.
- Bennett, G. K. (1944). *Test of productive thinking*. New York: Psychological Corporation.
- Biller, H. B., Singer, D. L., & Fullerton, M. (1969). Sex role development and creative potential in kindergarten-age boys. *Developmental Psychology*, 1, 291–296.
- Binet, A. (1911). *Les idées moderne sur les enfants [Modern ideas on children]*. Paris: Ernest Flammarion.
- Boden, M. A. (1990). *The creative mind: Myths and mechanisms*. London: Abacus.
- Calvin, W. (1998). *How brains think: Evolving intelligence, then & now*. London: Phoenix.
- Campione, J. C., Brown, A. L., & Ferrara, R. A. (1982). Mental retardation and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 392–492). New York: Cambridge University Press.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, 97, 404–431.
- Cattell, R. B. (1949). *Culture free intelligence test, Scale 1, handbook*. Champaign, Illinois: Institute of Personality and Ability.
- Cattell, R. B. (1967). The theory of fluid and crystallized general intelligence checked at the 5–6 year-old level. *British Journal of Educational Psychology*, 37, 209–224.
- Chi, M. T. H. (1997). Creativity: Shifting across ontological categories flexibly. In T. Ward, S. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 209–234). Washington, DC: American Psychological Association.
- Christensen, B. T. (2007). The relationship of analogical distance to analogical function and preinventive structure: The case of engineering design. *Memory & Cognition*, 35, 29–38.
- Colunga, E., & Smith, L. B. (2008). Flexibility and variability: Essential to human cognition and the study of human cognition. *New Ideas in Psychology*, 26, 158–192.
- Cramer, B. (1995). Infant creativity. *Infant Mental Health Journal*, 16, 21–27.
- Cronbach, L. J. (1968). Intelligence–creativity–parsimonious reinterpretation of Wallch–Kogan Data. *American Educational Research Journal*, 5, 491–511.
- Dehn, N., & Schank, R. (1982). Artificial and human intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 352–391). New York: Cambridge University Press.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1–26.
- Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, 68, 589–599.
- Eisenman, R., Platt, J. J., & Darbes, A. (1968). Creativity, intelligence, and achievement. *Psychological Reports*, 22, 749–754.
- Elias, J., & Dasgupta, S. (2005). A cognitive model of the engineering design mind. In J. S. Gero, & N. Bonnardel (Eds.), *Studying Designers '05* (pp. 101–116). University of Sydney, Australia: Key Centre of Design Computing and Cognition.
- Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research, and applications*. Cambridge, MA: MIT Press.
- Fuster, J. M. (2005). The cortical substrates of general intelligence. *Cortex*, 41, 228–229.
- Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1, 225–241.
- Gentner, D. (1983). Structure mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170.
- Getzels, J. W. (1975). Problem finding and the inventiveness of solutions. *Journal of Creative Behavior*, 9, 12–18.
- Getzels, J. W. (1987). Creativity, intelligence, and problem finding: Retrospect and prospect. In S. G. Isaksen (Ed.), *Frontiers of creativity research: Beyond the basics* (pp. 88–102). Buffalo, NY: Bearly Limited.
- Getzels, J. W., & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. New York: Wiley.
- Ghiselin, B. (Ed.). (1952/1985). *The creative process*. Los Angeles: University of California.
- Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science*, 16, 395–429.
- Guilford, J. P. (1950). Creativity. *American Psychologist*, 5, 444–454.
- Guilford, J. P. (1956). The structure of intellect model. *Psychological Bulletin*, 53, 267–293.
- Guilford, J. P. (1959). Three faces of intellect. *American Psychologist*, 14, 469–479.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Guilford, J. P. (1987). Creativity research: Past, present and future. In S. Isaksen (Ed.), *Frontiers of creativity research* (pp. 33–65). Buffalo, NY: Bearly.
- Guilford, J. P., Christensen, P. R., Merrifield, P. R., & Wilson, R. C. (1978). *Alternate uses: Manual of instructions and interpretation*. Orange, CA: Sheridan Psychological Services.
- Guthke, J. (1992). Learning tests: The concept, main research findings, problems and trends. *Learning and Individual Differences*, 4, 137–151.
- Guthke, J. (1999). Traditions and trends in intelligence testing. *Zeitschrift für Psychologie*, 207, 339–361.
- Haensly, P. A., & Reynolds, C. R. (1989). Creativity and intelligence. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of Creativity* (pp. 111–134). Plenum Press: New York.
- Halpern, D. F. (2003). Thinking critically about creative thinking. In M. A. Runco (Ed.), *Critical creative processes* (pp. 189–208). Cresskill, NJ: Hampton Press.
- Horn, R. (2009). *SPM-classic, -parallel, -plus manual* (pp. 27–29). Frankfurt am Main: Pearson.
- Indurkha, B. (1992). *Metaphor and cognition*. London: Kluwer Academic Press.
- Jaarsveld, S., Lachmann, T., Hamel, R., & van Leeuwen, C. (2010). Solving and creating Raven Progressive Matrices: Reasoning in well and ill defined problem spaces. *Creativity Research Journal*, 22, 304–319.
- Jaarsveld, S., & van Leeuwen, C. (2005). Sketches from a design process: Creative cognition inferred from intermediate products. *Cognitive Science*, 29, 79–101.
- Jung, R. E., Gasparovic, C., Chavez, R. S., Flores, R. A., Smith, S. M., Caprihan, A., et al. (2009). Biochemical support for the “threshold” theory of creativity: A magnetic resonance spectroscopy study. *Journal of Neuroscience*, 19, 5319–5325.
- Kaufmann, G. (2003). What to measure? A new look at the concept of creativity. *Scandinavian Journal of Educational Research*, 47, 235–251.
- Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education*, 16, 57–66.
- Koestler, A. (1964). *The act of creation*. London: Penguin.
- Kozbelt, A. (2006). Dynamic evaluation of Matisse's 1935 'Large Reclining Nude'. *Empirical Studies of the Arts*, 24, 119–137.
- Kozbelt, A. (2008). Hierarchical linear modeling of creative artists' problem solving behaviors. *The Journal of Creative Behavior*, 42(3), 181–200.
- Kratzmeier, H., & Horn, R. (1988). *Standard progressiver matrices, manual*. Weinheim, Germany: Beltz.
- Lau, S., & Cheung, P. C. (2010). Developmental trends of creativity: What twists of turn do boys and girls take at different grades? *Creativity Research Journal*, 22, 329–336.
- Liu, Y. C., Chakrabarti, A., & Bligh, T. (2002). Towards an 'ideal' approach for concept generation. *Design Studies*, 24, 341–355.
- Mackinnon, D. W. (1962). The nature and nurture of creative talent. *American Psychologist*, 17, 484–495.
- Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complex continuum in the radex and hierarchical models of intelligence. *Intelligence*, 7, 107–127.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, 69, 220–232.
- Miller, R. (1970). *Miller Analogies Test*. San Antonio: Pearson.
- Mullineaux, P. Y., & Dilalla, L. F. (2009). Preschool pretend play behaviors and early adolescent creativity. *Journal of Creative Behavior*, 43, 41–57.
- Mumford, M. D., Baughman, W. A., & Sager, C. E. (2003). Picking the right material: Cognitive processing skills and their role in creative thought. In Mark A. Runco (Ed.), *Critical creative processes* (pp. 19–68). Cresskill, NJ: Hampton Press.
- Mumford, M. D., Hunter, S. T., Eubanks, D. L., Bedell, K. E., & Murphy, S. T. (2007). Developing leaders for creative efforts: A domain-based approach to leadership development. *Human Resource Management Review*, 17, 402–417.



- Necka, E. (2003). Creative interaction: A conceptual schema for the process of producing ideas and judging the outcomes. In Mark A. Runco (Ed.), *Critical creative processes* (pp. 115–128). Cresskill, NJ: Hampton Press.
- Neubauer, A. (2010). Intelligenz und Kreativität aus Sicht der modernen Gehirnforschung. *Report Psychology*, 35, 322–332.
- Niaz, M., & De Nunez, G. S. (1991). The relation of mobility–fixity to creativity formal reasoning and intelligence. *Journal of Creative Behavior*, 25, 205–217.
- Parnes, S. J. (1972). *Creativity: Unlocking human potential*. Buffalo, NY: D. O. K. Publishers.
- Plucker, J. A., & Renzulli, J. S. (1999). Psychometric approaches to the study of human creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 35–61). Cambridge, UK: Cambridge University Press.
- Pomerantz, J. R., Sager, L. C., & Stoeber, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology. Human Perception and Performance*, 3, 422–435.
- Preckel, F., Holling, H., & Wiese, M. (2006). Relation of intelligence and creativity in gifted and nongifted students: An investigation of threshold theory. *Personality and Individual Differences*, 40, 159–170.
- Raven, J. C. (1938/1998a). *Standard Progressive Matrices, Sets A, B, C, D & E*. Oxford, England: Oxford Psychologists Press.
- Raven, J. C. (1976). *Advanced Progressive Matrices*. San Antonio, Texas: Harcourt.
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Section 3, Standard Progressive Matrices*. Oxford, England: Oxford Psychologists Press.
- Razumnikova, O. M. (2004). Gender differences in hemispheric organization during divergent thinking: an EEG investigation in human subjects. *Neuroscience Letters*, 362, 193–195.
- Runco, M. A. (1990). Implicit theories and ideational creativity. In M. Runco, & R. Albert (Eds.), *Theories of creativity* (pp. 234–254). Newbury Park, CA: Sage.
- Runco, M. A. (2003). Idea evaluation, divergent thinking, and creativity. In M. A. Runco (Ed.), *Critical creative processes* (pp. 69–94). Cresskill, NJ: Hampton Press.
- Runco, M. A. (2007). *Creativity. Theories and themes: Research, development, and practice*. Amsterdam, New York: Elsevier.
- Runco, M. A., & Albert, R. S. (1986). The threshold hypothesis regarding creativity and intelligence: An empirical test with gifted and non gifted children. *Creative Child and Adult Quarterly*, 11, 212–218.
- Sen, A. K. (1993). Correlations among creativity, intelligence, personality, and academic achievement. *Perceptual and Motor Skills*, 77, 497–498.
- Shavinina, L. V. (2001). Beyond IQ: A new perspective on the psychological assessment of intellectual abilities. *New Ideas in Psychology*, 19, 27–47.
- Silvia, P. J. (2008). Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan (1965). *Creativity Research Journal*, 20, 34–39.
- Silvia, P. J. (2008). Another look at creativity and intelligence: Exploring higher-order models and probable confounds. *Personality and Individual Differences*, 44, 1012–1021.
- Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence*, 4, 181–201.
- Simon, H. A., & Newell, A. (1971). Human problem solving: State of theory in 1970. *American Psychologist*, 26, 145–159.
- Simonton, D. K. (1990). Creativity in the later years: Optimistic prospects for achievement. *The Gerontologist*, 30, 626–631.
- Sligh, A. C., Connors, F. A., & Roskos-Ewoldsen, B. (2005). Relation of creativity to fluid and crystallized intelligence. *Journal of Creative Behavior*, 39, 123–136.
- Smilansky, J. (1984). Problem solving and the quality of invention: An empirical investigation. *Journal of Educational Psychology*, 76, 377–386.
- Smilansky, J., & Halberstadt, N. (1986). Inventors versus problem solvers: An empirical investigation. *Journal of Creative Behavior*, 20, 183–201.
- Smith, M., Ward, T. B., & Finke, R. A. (Eds.). (1995). *The creative cognition approach*. Cambridge, MA: Bradford.
- Spearman, C. (1904). 'General intelligence', objectively determined and measured. *The American Journal of Psychology*, 15, 201–293.
- Sternberg, R. J. (1982). Reasoning, problem solving, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 225–307). New York: Cambridge University Press.
- Sternberg, R. J. (1982). Conceptions of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 3–28). New York: Cambridge University Press.
- Sternberg, R. J. (2005). The WICS model of giftedness. In R. J. Sternberg, & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 237–243). (2nd ed.). New York: Cambridge University Press.
- Sternberg, R. J., & Lubart, T. E. (2003). The role of intelligence in creativity. In Mark A. Runco (Ed.), *Critical creative thinking* (pp. 153–188). Cresskill, NJ: Hampton Press.
- Szobiová, E. (2001). Relation of creativity and intelligence: Coincidence, similarities, and differences. *Ceskoslovenská Psychologie*, 45, 232–337.
- Torrance, E. P. (1966). *Torrance test of creative thinking*. Princeton, NJ: Personnel Press.
- Torrance, E. P. (1987). Teaching for creativity. In S. G. Isaksen (Ed.), *Frontiers of creativity research: Beyond the basics* (pp. 189–215). Buffalo, NY: Bearly.
- Unsworth, N., Redick, Th. S., Lakey, Ch. E., & Young, D. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, 38, 111–122.
- Urban, K. K. (2005). Assessing creativity: The Test for Creative Thinking–Drawing Production (TCT–DP). *International Education Journal*, 6, 272–280.
- Urban, K. K., & Jellen, H. G. (1995). *The Test of Creative Thinking–Drawing Production/Test zum Schöpferischen Denken–Zeichnerisch*. Lisse, Frankfurt: Swets Test Services.
- Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25, 1–28.
- Vernon, P. E. (Ed.). (1970). *Creativity*. London: Penguin.
- Viswanathan, V. K., & Linsey, J. S. (2009). Enhancing student innovation: Physical models in the idea generation process. *Frontiers of Education Conference, 2009. FIE '09. 39th IEEE*.
- Voss, J. F., & Post, T. A. (1988). On the solving of ill-structured problems. In M. T. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 261–286). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc..
- Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction*. New York: Holt, Rinehart, & Winston.
- Ward, W. C. (1968). Creativity in young children. *Child Development*, 39, 737–754.
- Ward, T. B. (2007). Creative cognition as a window on creativity. *Methods*, 42, 28–37.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. San Antonio: Psychological Corporation.
- Wertheimer, M. (1945/1968). *Productive thinking (Enlarged edition)*. London: Tavistock.
- Yamamoto, K. (1964). Threshold of intelligence in academic achievement of highly creative students. *The Journal of Experimental Education*, 32, 401–405.