This article was downloaded by: [Lachmann, Thomas]

On: 13 August 2010

Access details: Access Details: [subscription number 925555738]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Creativity Research Journal

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t775653635>

Solving and Creating Raven Progressive Matrices: Reasoning in Well- and Ill-Defined Problem Spaces

Saskia Jaarsveld^a; Thomas Lachmann^a; Ronald Hamel^b; Cees van Leeuwen^c

^a University of Kaiserslautern, Kaiserslautern, Germany ^b University of Amsterdam, The Netherlands ^c Brain Science Institute RIKEN, Wako-shi, Saitama, Japan

Online publication date: 10 August 2010

To cite this Article Jaarsveld, Saskia , Lachmann, Thomas , Hamel, Ronald and Leeuwen, Cees van(2010) 'Solving and Creating Raven Progressive Matrices: Reasoning in Well- and Ill-Defined Problem Spaces', Creativity Research Journal, 22: 3, 304 — 319

To link to this Article: DOI: 10.1080/10400419.2010.503541

URL: <http://dx.doi.org/10.1080/10400419.2010.503541>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Solving and Creating Raven Progressive Matrices: Reasoning in Well- and Ill-Defined Problem Spaces

Saskia Jaarsveld and Thomas Lachmann
University of Kaiserslautern, Kaiserslautern, Germany

Ronald Hamel
University of Amsterdam, The Netherlands

Cees van Leeuwen
Brain Science Institute RIKEN, Wako-shi, Saitama, Japan

We studied the development of creative cognition in children ranging from nursery school to Grade 6 (4–12 yr old, $N = 511$), performing a problem generation task. The task involved inventing a novel item for a classical problem solving task they had completed beforehand: the Raven Progressive Matrices (RPM). This task and the generating task both comprise matrixes of components, to which a set of transformational relations are applied; only in the first case these are inferred to solve a puzzle, but in the second they are invented to create one. We analyzed the matrixes invented in the generation task and compared them to those of the original solving task. We observed that (a) both in solving and generation, the ability to combine more than 1 relation increased with grade level, (b) within all 8 grades, except Grades 3 and 6, performance was uncorrelated between both tasks, (c) relations that were applied in the generation task often did feature in the solving task, and (d) relations occurring in both tasks were applied with different frequencies. Overall, we conclude that standard problem solving ability is not a precondition for creative reasoning and that the comprehension of relations between components featured in solving task differs from that applied in generation.

Everyday problems, such as “What do we serve for dinner?” or “How can I make up for lost time?” share an important characteristic with design problems in art and technology: they have no unique, correct solutions. Therefore, they are called *ill-defined problems*. How their solutions are appreciated will often depend on individual as well as on social and historical circumstances (Boden, 1990; Strohschneider & Guss, 1998).

The processes leading to solutions in such problems are referred to as *creative*, *productive*, or *generative*

reasoning. Here we consider creative reasoning to evolve, theoretically speaking, in abstract problem spaces that contain all possible steps between the current situation and its final solution (Anderson, 1983; Newell & Simon, 1972). The ill-defined character of these problems means that their problem spaces are, besides abstract, also indeterminate. This implies that valued, creative solutions could be obtained through novel interpretations about the rules that govern the steps reaching to a solution. Reinterpretation, without changing the nature of the problem, is permissible even in the initial situation (Chan & Fookkee, 1994; Runco, 2007). For instance, when experts are solving an architectural problem, they may add new definitions of form and content to the given problem (Goel & Pirolli, 1992). Reinterpretation typically involves uncovering implicit

We thank Frits Peters who wrote the software to compute the frequency profiles for *Rule*, *Components*, and *Specifications*.

Correspondence should be sent to Saskia Jaarsveld, Faculty of Social Sciences, Department of Psychology II, University of Kaiserslautern, Edwin-Schroedinger-Strasse 57, D-67663 Kaiserslautern, Germany. E-mail: jaarsvel@rhrk.uni-kl.de

requirements and making trade-offs among conflicting constraints (Yamamoto, Nakakoji, & Takada, 2000).

Standard, noncreative problem solving, by contrast, is understood in this framework as evolving in a determinate problem space, in which the rules that specify the steps, or moves allowed cannot be subject to reinterpretation (cf. a game of chess). Standard problem-solving tasks and generation tasks differ in this perspective, mainly according to the determinacy of the rules applied in their problem spaces.

Among the many studies on problem solving, relatively few have dealt with ill-defined problems (Akin & Akin, 1998; Getzels, 1975; Getzels & Csikszentmihalyi, 1976; Reiter-Palmon, Mumford, Boes, & Runco, 1997; Reiter-Palmon, Mumford, & Threlfall, 1998); some of these have included psychophysiological measurements (Fink et al., 2009). Fewer still have compared well- and ill-defined problem solving. Of these, some studies reported transfer from one problem type to the other (J. Jausovec, 2000; Kapa, 2007); others emphasized differences between well and ill defined problems, e.g., in processing demands (Schraw, Dunkle, & Bendixen, 1995); EEG and fMRI studies have further elaborated on these differences (N. Jausovec & Jausovec, 1997; Luo & Knoblich, 2007; Qiu et al., 2008) in efforts to localize the characteristic aspects of creative processes in distinct brain regions.

For meaningfully comparing well- and ill-defined problem solving, it is favorable to have, in both cases, the same knowledge domain, ensuring that, theoretically speaking, only the problem space differs between them. Unfortunately, with a few exceptions (Smilansky, 1984; Smilansky & Halberstadt, 1986), the important maxim of an identical knowledge domain is usually violated. Well-defined problem-solving tasks typically involve completing a logical structure according to well-defined rules (e.g., $2 + 2 = \dots$)—sometimes these rules are to be discovered, such as in pattern completion tasks. By contrast, typical ill-defined problem-solving tasks, insofar they use pattern completion, do not encourage application of a formalism. Rather they emphasize personalized experience (making a photo assembly of oneself) or involve practical knowledge of the world (inventing alternative uses for a brick). This is often considered to be the defining contrast between well- and ill-defined problem-solving processes although, in fact, it may have been imposed by the choice of tasks.

A similar type of contrast could be found between intelligence and creativity tests. Creativity tests can be cast as representing the ability of solving ill-defined problems. Some studies have investigated the relation of standard problem-solving tests and creativity tests (Barron, 1970; MacKinnon, 1962; Runco & Albert, 1986; Silvia, 2008a, 2008b; Sligh, Connors, & Roskos-Ewaldson, 2005; Sternberg, 1985; Torrance, 1962). But here, too, well-defined problem-solving tasks are

typically chosen to involve logical reasoning, as in the Raven Progressive Matrices test (RPM; Raven, Raven, & Court, 1998a, 1998b, 1998c, 1998d; for more information see Appendix A); whereas ill-defined problem solving tasks, such as the Test of Creative Thinking–Drawing Production (TCT-DP; Urban & Jellen, 1995), involve personalized associations. Therefore, when comparing performance on intelligence versus creativity tests we may consider that, in principle, performances reflect reasoning strategies which evolved from certain task domains. Thus, the resulting opposition of logic and creativity is possibly misleading.

This study aimed to compare the abilities of solving well- and ill-defined problems, by using the same knowledge domain. We adopted the RPM test as a standard problem-solving task. The RPM presents individuals with a series of items. Each item consists either of a continuous pattern in a large rectangle or a pattern divided into a 2×2 or 3×3 matrix. The cells of a matrix may contain a simple component (e.g., triangle, circle), or components made out of several subcomponents (e.g., two triangles placed side by side or combined to make a star of David). The cell in the bottom right corner of the matrix is left vacant for completion. The component that should fit into this cell has to be determined by the individual from relations such as: sameness, increment, or combination, which can be inferred from the components given. Below the matrix, a set of response alternatives is presented, from which the correct completion has to be chosen.

The RPM items represent well-defined problems: The knowledge necessary to infer a relation is specified by the problem itself, and the completion is unambiguous. Following Smilansky (1984) and Smilansky and Halberstadt (1986), our *generation* task required individuals to create a new RPM item. Thus, the generation task represents an ill-defined task in which a well-defined problem is to be created. Using this task secured that the knowledge domain in the well and the ill-defined task was the same. Both require the comprehension of a set of relations and reasoning about possible solutions.

Whereas RPM solving involves recognition of components and relations, in the ill-defined task they have to be created. This creation may be informed, for instance, by personalized associations (Karmiloff-Smith, 1990), which are able to enter the creative process because of the indeterminate character of the space in which it evolves. However, this does not have to be the case, because the domains are identical, components and their relations might simply be copied from the solving task. Therefore, if differences arise between both tasks in the components and their relations, they are likely to be the result of differences in reasoning strategies, rather than domain preferences.

Can we expect such differences? One hypothesis is that reasoning applied in well- and ill-defined problems is largely the same. This hypothesis relates to the theory

that new ideas are created based on prior acquired knowledge (Cacciari, Levorato & Cigogna, 1997; Ward, 1995; Weisberg, 1993; Wilkening, Schwarzer, & Rümmele, 1997). The level of comprehension reached in the solving task will be reflected in the creation task. Thus, we would expect that the rules applied in the generation task reflect those featured and solved in the RPM. Whereas we cannot test this hypothesis on individual problem solutions, we can, however, predict according to this hypothesis that the frequency distributions of rules and (sub)components of the creation task will be identical to those of the RPM solving task. In other words, any given rule would occur with equal frequency between both tasks. According to this hypothesis, we also would expect the pictorial characteristics of the (sub)components featuring in the RPM items to be reflected in the created ones.

Smilansky (1984) observed that, within a group of students, the combination of low RPM scores and high scores in the generation task failed to occur. From this, he concluded that creativity builds on established intelligence. If so, we might expect the rules used in creative tasks to reflect those in solving tasks, but with some delay needed for the shift from rule comprehension to its application in production. Therefore, we compared problem solving and creation across school levels. Comparison across school levels of problem creation, finally, allowed us to observe whether more advanced pupils have a higher developed ability to process complex information (Halford, 1993).

On the other hand, there are reasons to assume that the rules may differ more radically between problem solving and generation. We recall that in well-defined problems the rules are fixed, and in ill-defined problems they are not. This could give rise to qualitative differences between the tasks, as reinterpretation may lead to the emergence of rules not previously observed in the solving task. If such differences arise, we may conclude that classical and creative reasoning strategies differ, and are potentially independent faculties.

In addition, solving a standard problem does not always mean the problem is understood: Often a correct solution is accompanied by an incorrect line of verbal reasoning or is obtained without any conceptual understanding (Chi & VanLehn, 1991; Karmiloff-Smith, 1992; Mestre, 2002; Pine & Messer, 1999). Such distortions are less likely with illdefined problems, because their solutions characteristically evolve in attempts to explain the problem to oneself.¹ Because of this, we may expect

¹Because of this, ill-defined problems constitute a powerful mechanism for promoting understanding and conceptual advancement (Chi & VanLehn, 1991; Mestre, 2002; Siegler, 2005). They, therefore, are most suitable to assess a pupil's, student's, or job candidate's understanding, reasoning mechanisms, and presuppositions (Carey & Flower, 1989; Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001).

the ill-defined problems to reveal a quantitative shift in the frequency with which certain rules appear.

METHOD

Participants

There was a total of 511 children (264 girls = 52%), aged between 4 and 12 years old, from nursery ($N=64$) and elementary schools ($N=447$) situated in small towns in the west of the province South Holland, the Netherlands. We distinguished children according to nursery school levels (*younger*: between four and five years of age vs. *older*: between five and six years of age) and school grades (Grade 1 to Grade 6; for details see Table 1). School authorities gave approval to this study, and parents or caretakers signed a letter of consent before a child could participate in the study. They and the children were assured that individual results would not be reported back to the schools. The children received small presents for their cooperation.

Material

A standard problem-solving task and a problem-generation task were used to directly compare the cognitive abilities in dealing with well- and ill-defined problems within the same knowledge domain. The RPM were applied as problem solving task; the *generation* task required individuals to create a new RPM item.

RPM. The RPM test is a nonverbal standardized test to assess a person's capacity for coherent and clear

TABLE 1
Number of Participants and Average Age per School Grade

	Total Coloured Progressive Matrices ($N=206$)	Grade	N	M Age
Nursery school	64	younger	33	4.64
		older	31	5.68
Elementary school	447	1	41	6.73
		2	42	7.79
		3	59	8.81
	Standard Progressive Matrices ($N=305$)	4	132	9.80
		5	91	10.87
		6	82	11.91
Total	511			

Note. A total of 511 Dutch children participated in the study (52% girls); 64 children from nursery schools and 447 children of elementary schools. Age limits within grades were not absolute, but the averaged age increased with 1 year.

thinking. The RPM test is construed in such a manner that individuals learn to solve more difficult problems as they proceed through the items of the task. To assure a gradual increase in difficulty, the order of items in the Raven test is based on the Item Characteristic Curves. As the items get increasingly difficult, a critical item will occur, above which the individual is expected no longer to solve any more items correctly. Nevertheless, each individual completes the entire booklet. The number of items solved correctly constitutes the individual score.

The term *RPM* is commonly used to name a series of three tests corresponding to different subpopulations regarding the level of cognitive development (see Appendix A for a detailed description). The Coloured Progressive Matrices (CPM, 36 items) test was designed to assess the cognitive abilities of young children, mentally retarded, and elderly persons. The Standard Progressive Matrices (SPM, 60 items) test was designed for use with children as well as adults. The Advanced Progressive Matrices (APM) test, which was constructed for adults with above-average intellectual abilities, was not included in this study. Each test is contained in a booklet, which displays one incomplete matrix per page, together with a multiple choice of completion alternatives. A separate answering sheet is offered, on which individuals mark the alternative they consider to be the proper completion. Smilansky (1984) used only a part of the SPM (Series D and E). We used the complete series (A to E) and additionally applied the CPM because our study involved children from nursery and elementary schools.

To illustrate the principle of RPM items, Figure 1 shows a couple of hypothetical examples, both consisting

FIGURE 1 Raven progressive matrices test is construed in such a manner that participants learn (by example) as they proceed through the items of the task. Figure 1 shows hypothetical examples of matrices; in Figure 1a the components feature in a sequence that is identical in each row. However, in Figure 1b the sequence is identical to that of in Figure 1a, but it starts with a different component in each following row. The knowledge applied in Figure a together with the information from Figure 1b may allow the participant to infer the relation featured in the latter matrix. As the items get more difficult, at a certain point the participant can no longer identify the correct relation.

of a 3×3 matrix. In the first example (Figure 1a), the uppermost two rows contain a sequence of cube, cross, and circle; the last row contains only the first two components, and therefore the correct completion would be the circle. This example embodies a rule we shall call *change*, by which components, for instance, are in the same order across rows, and remain identical within columns. In the other matrix (Figure 1b) the first two rows both contain a cube, a cross, and a circle, but in a different order. The last row contains two of these components (cross and circle), again in a different order. From this it follows that the completion should be the square. Here the example embodies a rule that we call *succession*, by which, for instance, the order of the components across rows, and their identity across columns should differ. The knowledge applied to solve the first item, together with the presentation of the second item will allow the individual to infer (Raven used the term “*educe*”) this second rule (“the assessment of the *eductive* component obtainable with the *Progressive Matrices*.” Raven, Raven, & Court, 1998e, p. 1). Note that rules can apply in horizontal or in vertical direction.

Creative reasoning task. The creative reasoning task (CRT) consists of an empty outline of a RPM item printed on a form (see Figure 2). Children are asked to invent their own Raven matrix and draw it within the outline of the form. The bottom right cell of the matrix, which is highlighted on the form, should be left vacant. The CRT corresponding to the CPM requires children

FIGURE 2 The form used in the creative reasoning task (CRT). Children were asked, after having made the Colored Progressive Matrices (CPM) or Standard Progressive Matrices (SPM), to create a Raven-like item. The large square is for the matrix, the smaller boxes for the alternatives (six in the CPM and eight in the SPM condition). Children were asked to drawn their “correct” solution in the alternatives cell number five.

to draw six completion alternatives; the CRT corresponding to the SPM requires eight alternatives. These alternatives should be drawn in the empty cells on the bottom of the page. One and only one of the alternatives solutions should be correct. This solution should be drawn in the alternatives cell number five.

Procedure

The session started with a general introduction to the team of experimental assistants and to the procedure, and then an introduction to the RPM task was given according to standard procedures (Raven et al., 1998a). All children first performed the RPM test. Assistants handed out the booklets, forms, and pens during the introduction, and collected the response forms after the children had completed the tests. Children of nursery school and elementary school Grades 1 to 3 performed the CPM; children from Grade 4 to 6 performed the SPM (see Table 1). After completion of the RPM (either CPM or SPM), children were instructed to create an RPM item. It was stressed that they should try to make this item as difficult as they possibly could so that it would be a hard puzzle for others to solve. A session comprised about 65 min; about 50 min were destined for the RPM test and about 15 min for the CRT.

Nursery school children performed the tasks individually, each under supervision of an assistant in a separate room. Elementary school children performed the tasks in groups at their own classroom desks, which were arranged in identical alignment and placed at equal distances from each other. There were 28 to 31 children in a classroom. They were asked to keep reading material at hand in case they completed the task earlier than others and had to wait until the sessions had elapsed.

Analysis

The scores of the CPM were converted to SPM scores according to the scale provided by Raven et al. (1998c) to enable direct comparisons between grade levels. The scores, as such, cannot yield an insight into the knowledge that was applied in the performance of the tasks. To observe such differences, we must score the components and relations of the RPM items and compare them to ones created in the generation task. Smilansky (1984) scored the items created for the number of components (which he called "elements" and, unfortunately, did not define in detail, p. 380) and the number of relations among these. He correlated this measure with the individual scores of the RPM. We endorse the general strategy to compare RPM scores with the production of a Raven item, but chose a different strategy for our analyses. This was because the score of a created item,

as applied in Smilansky (1984), does not provide us with information about the type of rules used in matrix completion. As a result, this scoring method fails to connect with the notion of problem solving as operating in a problem space. Our method was focused on comparing the rules of completion applied in solving a generation tasks.

We scored the performance on the RPM and CRT according to the following measures, which are explained in the following sections: *rule*, *component*, and *specification*. Basically, the greater was the complexity in *rule* and the higher was the variety in *component* and *specification*, the higher was the score of the creative reasoning task. Complexity was defined as the ability to process more than one rule at once (English, 1998; Halford, 1993). *Rule* represents mainly convergent thinking, and *component* and *specification* represent mainly divergent thinking. Generally, however, both types of thinking are required in a creative design process (Cropley, 2006; Jaarsveld & van Leeuwen, 2005) and, thus, were assumed to apply in particular in the present creative reasoning task.

Rules

The score for *rule* indicates the number of relations between the components of a matrix, which represents its complexity. Each rule correctly applied either in *horizontal* (H) or *vertical* (V) direction increments the score by 1. For instance, when components change along the rows (e.g., square, triangle, and circle), the rule change (H) was applied; when these components also change their texture along the column (e.g., open, shaded, and filled), this was scored with the rule change (V). Assuming both rules were correctly applied, the resulting score for such an item would be *rule* = 2.

In these scores, we took into account the fact that the relevance of rules may depend on the response alternatives. For instance, in the aforementioned example, the texture change is relevant only, strictly speaking, if the response alternatives contain both a filled and a differently textured circle. However, we scored for two points the correct solution of a filled circle in absence of a differently textured one, as long as there were differently textured components among the response alternatives, from which we could infer that texture was a relevant variation.

A total of 12 rules were distinguished, of which five (Rules 5 to 9, see the following) were based on the ones that Carpenter, Just, and Shell (1990) defined for the SPM. We defined three additional rules for the CPM (Rules 2 to 4, see the following) and an additional one for the SPM (Rule 11). Three more rules that were neither featured in the SPM nor in the CPM, were observed to occur in the CRT (Rules 1, 10, and 12,

FIGURE 3 Raven-like items generated in the CRT. Participant number 179, 5 years old, represents rule 1, *idiosyncratic and semantic coherence*. Participant number 36, 10 years old, represents rule 3, *continuous pattern*. Participant number 64, 12 years old, represents rule 12, *groups of three components*. Participant number 354, 10 years old, represents rule 2, *four identical components* and rule 4, *symmetry*. Participant number 56, 12 years old, represents rule 5, *change*, rule 6, *increase and decrease*, and rule 8, *rotation and succession*. Participant number 59, 12 years old, represents rule 7, *exchange and combine*.

see the following). Rules are explained in the following, illustrated by examples of created items taken from the present study (see Figure 3).

Rule 1: Idiosyncratic and semantic coherence. This rule applies when a relation between components follows a personalized theme or scene, for instance the last birthday, Christmas, or a sunny day, as in the example displayed in Figure 3 (Participant 179, 5 years old) or something known only to the inventor.

Rule 2: Four identical components. This rule applies when, in a 2×2 matrix, four identical

components are linked as pairs or as a fourfold without any rule, except identity in form and texture.

Rule 3: Continuous pattern. This rule applies when components feature in a painting or a wallpaper kind of pattern instead of in discrete cells; solving the problem amounts to pattern interpolation (see Figure 3, Participants 36, 10 years old).

Rule 4: Symmetry. This rule applies when parts of an item are rendered in a symmetrical or mirror image way (see Figure 3, Participant 354, 10 years old).

Rule 5: Change. This rule equals Carpenter, Just, and Shell's (1990, p. 408) *constant in a row*, which applies when "the same value occurs throughout a row, but changes down a column" (see Figure 3, eye, hand, mouth; Participant 56, 12 years old).

Rule 6: Increase and decrease. This rule applies when components, subcomponents, or attributes of components are subject to increase or decrease, either in size, number (Figure 3, Participant 64, 12 years old) or any other dimension. This rule is a modified version of Carpenter et al.'s (1990) *quantitative pair-wise progression*. These authors included, besides the operation of increase/decrease, that of rotation. We considered this to be a separate rule (Rule 8).

Rule 7: Exchange and combine. Exchange applies when components or subcomponents shift location gradually across three cells. Combine applies when in any constellation two components of a row or a column combine to form the subcomponents of a third one (see Figure 3, Participant 59, 12 years old). Combine corresponds to Carpenter et al.'s (1990) *figure addition and subtraction*.

Rule 8: Rotation and succession. Rotation applies when components or subcomponents in a row or a column differ systematically in orientation or when the cells in a row or a column differ in the location of the subcomponents in a manner that reflects a rotation of the component. The shift from the first to second row is to be repeated in the shift from second to third row, and this rule is to determine the missing component. Succession applies when a component or subcomponent appears once in the first, once in the second and once in the third row, but never appears in the same cells (see Figure 3, Participant 56, 12 years old; frame and egg). The rule *rotation and succession* is a combination of Carpenter et al.'s (1990) *quantitative pair-wise progression* and *distribution of three values*.

Rule 9: Disappear and remain. *Disappear* applies when a subcomponent or attribute that is shared by two components disappears, and ones not shared by both remain the same, and are combined into the third component. *Remain* applies when a subcomponent or attribute that is shared by two components features in the third and the one that is not shared disappears. In this case, the shared subcomponent or attribute features in all three components. *Disappear* is identical to Carpenter et al.'s (1990) *distribution of two values* and features in the SPM. *Remain* features only in the APM version and therefore was not defined by Carpenter and colleagues.

Rule 10: Indication of form, texture, number, or orientation. Rule 10 applies when the second component specifies the transformation that must be performed on the first one of the same row or column to produce the third element. The manipulations indicated in the first two rows or columns can together apply to the third one. Transformations can involve shape, texture, orientation, etc. This rule is thus reserved for transformations less obvious than increase or decrease in number, subtraction or addition, or rotation.

Rule 11: Indication of arithmetic operation. This rule applies when subcomponents or attributes of the second component specify the mathematical operation that is performed on the first one to construct the third one. For instance, one attribute of the second component, say: a white dot, may symbolize increment; meaning that the three strokes in the first component are raised to four in the third component, and another, say: a black dot, may indicate decrement; from three strokes in the first to two strokes in the third component.

Rule 12: Groups of three components. This rule applies when components are equally distributed over three categories (e.g., three vehicles, three animals, and three circles) but are irregularly dispersed over the matrix (see Figure 3, Participant 64, 12 years old). The components do not appear with the regularity that is mandatory for rule 8, *rotation* and *succession*.

Components

Smilansky's (1984) score of the number of different elements included within the design gives some indication of fluency, but does not provide further information about the pictorial characteristics of the design. Our *component* score refers to the number of different

subcomponent types in the completed matrix. A component may consist of one or several subcomponents (see Figure 3, Participant 59); for a drawing, components as in Figure 3, Participant 179, the sun and the flower are counted separately; for a continuous pattern, as in Figure 3, Participant 36, the roots, the stem, and the branches are counted as subcomponents. To distinguish subcomponents from each other, we used the criterion that they should belong to distinct semantic or pictorial categories. This distinguished figurative subcomponents such as trees, flowers, etc., predominant in young children but also abstract geometrical shapes, such as triangles or squares. The subcomponents could overlap in space, for instance a cross superimposed on a circle. They should be distinguished from attributes, such as the shading of a component. We used the criterion of separability (Garner, 1970; Pomerantz, Sager, & Stoeve, 1977); cross and circle can exist as separate semantic or pictorial elements, whereas circle and shading cannot. The geometrical components identified according to these criteria were scored in nine different categories: *line*, *square*, *circle*, *triangle*, *diamond*, *cross*, *beam*, *oval*, and *dot*. The *figurative* category is, in principle, a boundless category, and therefore did not contain any fixed set of subcategories.

Specifications

Specification scored the occurrence of different attributes or pictorial characteristics which define a given (sub)component. Seven categories of *specification* were distinguished. The first three categories, *black-white*, *shaded*, and *speckled*, applied to surface texture, the fourth category, *undulated*, applied when any of the subcomponents were rendered with an undulated outline. The fifth category, *size*, applied to the rendering of any component or subcomponent, in a smaller or larger size. The sixth category, *orientation*, applied to the rendering of components or subcomponents in three different orientations, irrespective of the type of (sub)component affected. The seventh category, *number*, applied when a subcomponent was rendered more than once, either within a component, or across the whole item. The three nontexture categories, *size*, *orientation*, and *number*, applied only as long as they were ornamental and did not constitute a rule. Here, too, the interrater reliability offered a check on arbitrariness.

As a check on the reliability of our scoring method, multiple ratings were obtained, and interrater correlations were calculated. Using the categories of *rule*, *component*, and *specification*, we compared the scores of items featured in the RPM, items participants were able to solve, and items created in the CRT. We obtained the cumulative frequency profiles for *rule*,

component, and *specification* up to the last item a participant solved correctly.

RESULTS AND DISCUSSION

Interrater reliability was calculated separately for *rule*, *component*, and *specification* on two different subsets of the data ($N_1 = 95$, $N_2 = 29$). The results ranged from $\kappa = 0.95$ (for specifications) to $\kappa = 0.93$ (for rules; see Siegel & Castellan, 1988). These results indicate a strong correspondence between different raters. Table 2a provides for each grade the percentage of the total number of rules contained in items featured in the RPM test, Table 2b displays the same for items solved, and

Table 2c for items created in the CRT. These tables reveal both qualitative and quantitative effects.

Qualitative Effects

Rules. We deal mostly with the issue of whether the rules in the items created in the CRT reflected those featured and solved. Note that the items featured differed between, on the one hand, nursery school and Grades 1 to 3, who received the CPM; and, on the other hand, the higher grade levels (Grades 4 to 6) who received the SPM (see Table 2a).

According to Table 2b, for children who received the CPM, we observed that the younger nursery school children could solve items containing rules 2, 3, and 4,

TABLE 2
Percentages of the Rules Featured and Solved in the Coloured Progressive Matrices (CPM) or Standard Progressive Matrices Test (SPM), and Applied in the Creative Reasoning task (CRT)

Test	N	School Grade	Rules Featured											
			1	2	3	4	5	6	7	8	9	10	11	12
CPM	33	Younger nursery	—	30.36	21.43	25.00	23.21	—	—	—	—	—	—	—
CPM	31	Older nursery	—	30.36	21.43	25.00	23.21	—	—	—	—	—	—	—
CPM	41	1	—	30.36	21.43	25.00	23.21	—	—	—	—	—	—	—
CPM	42	2	—	30.36	21.43	25.00	23.21	—	—	—	—	—	—	—
CPM	59	3	—	30.36	21.43	25.00	23.21	—	—	—	—	—	—	—
SPM	132	4	—	6.31	10.81	8.11	20.72	18.92	10.81	13.51	10.81	—	1.80	—
SPM	91	5	—	6.31	10.81	8.11	20.72	18.92	10.81	13.51	10.81	—	1.80	—
SPM	82	6	—	6.31	10.81	8.11	20.72	18.92	10.81	13.51	10.81	—	1.80	—

Test	N	School Grade	Rules Solved											
			1	2	3	4	5	6	7	8	9	10	11	12
CPM	33	Younger nursery	—	39.64	41.98	18.38	0.00	—	—	—	—	—	—	—
CPM	31	Older nursery	—	39.31	35.00	24.86	0.83	—	—	—	—	—	—	—
CPM	41	1	—	33.31	23.57	25.64	2.68	—	—	—	—	—	—	—
CPM	42	2	—	31.64	22.21	24.64	4.42	—	—	—	—	—	—	—
CPM	59	3	—	27.21	17.89	22.17	6.72	—	—	—	—	—	—	—
SPM	132	4	—	8.52	15.84	11.22	23.69	19.48	10.01	9.71	1.53	—	—	—
SPM	91	5	—	8.17	15.17	10.78	23.96	19.38	10.65	10.12	1.77	—	—	—
SPM	82	6	—	7.67	14.33	10.22	23.73	18.96	11.16	10.73	3.20	—	—	—

Test	N	School Grade	Rules Applied											
			1	2	3	4	5	6	7	8	9	10	11	12
CPM	33	Younger nursery	76.92	7.69	15.38	0	0	0	0	0	0	0	0	0
CPM	31	Older nursery	51.85	7.41	40.74	0	0	0	0	0	0	0	0	0
CPM	41	1	5.71	94.29	0	0	0	0	0	0	0	0	0	0
CPM	42	2	7.50	2.50	87.50	0	0	2.50	0	0	0	0	0	0
CPM	59	3	7.02	19.30	70.18	3.51	0	0	0	0	0	0	0	0
SPM	132	4	2.63	1.75	53.51	1.75	11.40	4.39	2.63	15.79	0	0	0	6.14
SPM	91	5	1.32	3.95	67.11	1.32	3.95	6.58	3.95	7.89	0	0	0	3.95
SPM	82	6	0	3.85	51.28	1.28	5.13	8.97	8.97	12.82	0	1.28	1.28	5.13

Note. Values represent relative frequencies of rules featured and solved in one of the RPM tests, and applied in the CRT.

"Rule 1 = Idiosyncratic and Semantic Coherence, Rule 2 = Four identical Components, Rule 3 = Continuous Pattern, Rule 4 = Symmetry, Rule 5 = Change, Rule 6 = Increase and Decrease, Rule 7 = Exchange and Combine, Rule 8 = Rotation and Succession, Rule 9 = Disappear and Remain, Rule 10 = Indication of Form, Texture, Amount or Orientation, Rule 11 = Indication of Arithmetic Operation, Rule 12 = Groups of Three Components.

but not rule 5, *change*. Older nursery school children could solve these items incidentally, and the first and second graders did so with increasing frequency. However, this rule remained the most difficult one in the CPM. Children receiving the SPM were able to solve items featuring any of the rules, except rule 11, *indication of arithmetic operation*.

As shown in Table 2c, the preference for certain rules in the CRT, may indicate lines along which creative reasoning develops. Younger nursery school children preferred rule 1, *idiosyncratic and semantic coherence*. As the example of Figure 3 (Participant 179) suggests, this may express a story or anecdote somehow relevant to the individual biography of the child. It is noteworthy that these participants, who made the CPM, hardly used rule 2, *four identical components*, although it constitutes a considerable part of this test (Table 2a) and was frequently solved (Table 2b).

Older nursery school children preferred rule 1 as well as rule 3, *continuous pattern*. As Figure 3 (Participant 36) illustrates, children applying this rule often deleted a piece in an overall pattern, figurative or non figurative.

Notably, neither of the nursery school groups preferred rule 2, *four identical components*, even though it was the one featured most frequently in the CPM and was also frequently solved. By contrast, Grade 1 showed a strong preference for rule 2, which was applied almost exclusively in the CRT, in disregard of the other rules featured and solved with high frequency. As Figure 3 (Participant 354) illustrates, children applying this rule often made use of several subcomponents, which are pair wise related by means of repetition or a continuous pattern at the level of the subcomponent.

The Grades 2 to 6 showed a strong preference for rule 3, *continuous pattern*. This preference remained invariably strong, even though only the Grades 2 and 3 had frequently engaged with it through the CPM. It is far less frequent in the SPM, which was used for Grades 4 to 6. On the other hand, rule 5, *change*, which children receiving the CPM found difficult to solve, did not occur at all in the CRT of these children. For children receiving the SPM, where this rule was the most frequent one, it was still topped in the CRT by rule 3. The rules containing the manipulations *increase*, *combination* and *rotation*, the rules 6 to 8, respectively, were equally applied in small frequencies.

Finally, rule 1, *idiosyncratic and semantic coherence*, a rule which did not feature at all, in neither the CPM nor the SPM, was most predominant in all nursery children and continued to occur throughout elementary school, with the highest grade as the only exception.

Components. The percentage of participants who created figurative items decreased with increasing grade level, $r_s = -.671$, $n = 8$, $p < .05$ (r_s or p [rho]), is

Spearman's rank correlation coefficient, a nonparametric measure of statistical dependency between two variables, in this case between percentage of figurative items and grade level). Of the younger nursery school children, 58% created an item of a figurative character, while in Grade 6 figurative items were down to 21%. The coincidence with figurative components and rule 1 occurred only in nursery school, as these children applied this rule most frequently (younger nursery school children, $n = 20$, 85%; older nursery school children, $n = 14$, 79%).

Nursery school children most frequently applied the square and the circle, followed by the line. By contrast, the children from Grades 1 to 6 applied most frequently the line, followed by the circle and the square. Least applied in all children were the oval and diamond. This could indicate that the intricate structure of these components is a challenge to the hand-eye coordination of children in nursery and elementary schools.

Specification. The younger nursery school children applied most frequently the specification *black-white*, followed by *shaded*. Least applied in all children was *speckled*, probably due to the perseverance necessary to draw this type of texture. The preference for *black-white* decreased with grade, $r_s = -.786$, $n = 8$, $p < .01$. All non-texture categories increased with grade level, for *size*, $r_s = .802$, $n = 8$, $p < .01$, for *orientation*, $r_s = .886$, $n = 8$, $p < .01$, and for *number*, $r_s = .731$, $n = 8$, $p < .05$ (all correlations one-sided). This indicates that components are increasingly drawn in variable numbers, and different sizes and orientations.

Quantitative Effects

Total Scores

To obtain an overall score for the CRT we added the scores for *rule*, *component*, and *specification*. This score, as well as the score for the CPM and RPM, are displayed for each Grade level in Table 3 (CPM converted to SPM scores, as previously discussed). The CRT score was found to correlate significantly with Grade, $r_s = .667$, $n = 8$, $p < .05$, as did the score of the solving task, $r_s = 1.00$, $n = 8$, $p < .01$ ($r = .99$, $p = 0.01$; see the scores of the solving task, SPM, in Table 3). The SPM scores did not differ between those who created a nonfigurative and those who created a figurative item on the CRT for any of the grades except for Grade 6. Here, the RPM scores of those who created figurative components ($M = 44.18$, $SD = 6.54$, $n = 17$) were lower than those who created non figurative components ($M = 46.02$, $SD = 5.75$, $n = 65$), $t(82) = 1.903$, $p < .05$ (one sided). Only two of the eight grades showed a correlation between the solving task and the creative reasoning

TABLE 3
Test Scores of the Solving and Generation Task

Grade	SPM ^a	CRT ^b	SPM scores	
			Nonfigurative ^c	Figurative ^d
Younger nursery school	14	3.85	14	14
Older nursery school	17	4.52	17	18
1	23	3.46	23	23
2	25	3.83	25	25
3	30	4.12*	30	30
4	40	5.23	40	39
5	43	4.91	43	43
6	46	5.12*	46	44*

Note. Scores of the solving (SPM) and generation task (CRT) increased with grade. However, within the eight groups no relation was found between the scores of both tasks, except for Grade 3 and Grade 6.

^aScores include CPM scores converted to SPM standards. Scores are the number of items solved correctly. ^bCRT scores are additions of the mean frequencies of rules, components, and specifications. ^{c/d}SPM mean scores related to the graphical character of the components in the CRT. These scores only differed in Grade 6.

* $p < .05$.

task: Grade 3, $r = .28$, $T = 2.20$, $df = 57$, $p < .05$, and Grade 6, $r = .25$, $T = 2.31$, $df = 80$, $p < .05$ (T is a statistic of the estimated difference between the two samples based on Student's t distribution, as the population standard deviation is unknown). From the absence of a correlation in the remaining six grades we may infer that scores on the solving task will fail to predict performance on the creative reasoning task.

Rules, Components, and Specifications

Frequencies of rules, components, and specifications of items featured or solved depended on whether participants received a CPM ($N = 206$) or SPM ($N = 305$). For both groups, we calculated chi squares of independence over the relative frequency distributions. In distributions with $df = 1$ and expectancy values lower than five, we corrected for continuity, and, additionally, performed Fisher's exact test to double check the significance of the effects. In distributions with more cells, the categories that showed expectancies lower than five were excluded from the chi square calculations (cf. Siegel & Castellan, 1988).

Rules. The number of items created in the CRT that could not be scored with a correctly applied rule, was not related to grade level. Numbers ranged from 21% for the younger nursery school children, to 15% for Grade 6, with a minimum of 5% for Grades 2 (see Table 4). The rule score increased with grade level,

TABLE 4
Rules, Components, and Specifications Applied in the CRT

Grade	Numbers of Rules ^{a**}					CS ^{b*}
	0	1	2	3	4	
Younger nursery	21	79	0	0	0	3.06
Older nursery	13	87	0	0	0	3.65
1	15	85	0	0	0	2.61
2	5	93	2	0	0	2.86
3	7	90	3	0	0	3.15
4	15	67	14	3	0	4.17
5	18	74	7	1	1	3.97
6	15	65	16	2	2	3.99

Note. With increasing grade level items featured more rules (Number of rules = 1 to 4), and components and specifications (CS), indicating an increase in the complexity of the items created in the CRT.

^aPercentage of the number of rules applied within a created item.

^bSum of the averaged frequencies of components and specifications. r_s ** $p < .01$, * $p < .05$.

$r_s = .826$, $n = 8$, $p < .01$, as did the number of rules applied within one item, $r_s = .964$, $n = 8$, $p < .01$ (see Table 4), both indicating an increase in complexity of generation.

For children who received the CPM, the distribution of rules featured and solved (see Table 2) differed, $\chi^2(8) = 37.38$, $p < .001$ (calculated over rules 2 to 4). This difference confirms the observation that some rules, as identified in the qualitative analysis, were more difficult than others. As expected, the frequency distributions of rules featured and applied differed also, $\chi^2(4) = 553.64$, $p < .001$ (calculated over rules 2 and 3), as did the distribution of rules solved and applied, $\chi^2(4) = 578.45$, $p < .001$ (calculated over rules 2 and 3).

For those who performed the SPM, the distribution of the rules featured and solved did not differ. This indicates that all rules were approximately equally difficult for this group. As expected, the distribution of rules featured and applied, $\chi^2(8) = 680.84$, $p < .001$ differed, as did the distribution of rules solved and applied, $\chi^2(8) = 440.62$, $p < .001$ (both calculations over rule 3 and rules 5 to 8). We may conclude that, overall, knowledge applied in solving differed from that applied in creation.

Table 5 deals with the question whether this conclusion applies to all grades. According to the table, the frequencies with which rules were applied in items created differed from those in the items solved or featured within all the grade levels. Hence, we may conclude that the divergence between reasoning in problem solving and problem generation is not a developmental trend, but exists from early school days on.

It may be argued that generating is more difficult than solving (Smilansky, 1984) and perhaps generation

TABLE 5
Comparisons of Rules Featured and Solved in the RPM, and Applied in the CRT

<i>N</i>	<i>School Grade</i>	<i>Solved-Featured^a</i>	<i>Applied-Featured^b</i>	<i>Applied-Solved^c</i>
33	Younger nursery	$\chi^2(2) = 24.31, p < .001$	$\chi^2(1) = 18.63, p < .001$	$\chi^2(1) = 42.60, p < .001$
31	Older nursery	$\chi^2(2) = 11.23, p < .01$	$\chi^2(1) = 34.75, p < .001$	$\chi^2(1) = 26.83, p < .001$
41	1	$\chi^2(2) = 0.52, p > .05$	$\chi^2(1) = 156.05, p < .001$	$\chi^2(1) = 135.17, p < .001$
42	2	$\chi^2(2) = 0.09, p > .05$	$\chi^2(1) = 229.28, p < .001$	$\chi^2(1) = 218.77, p < .001$
59	3	$\chi^2(2) = 1.23, p > .05$	$\chi^2(1) = 114.92, p < .001$	$\chi^2(1) = 155.08, p < .001$
132	4	$\chi^2(6)^d = 5.88, p > .05$	$\chi^2(4)^e = 190.56, p < .001$	$\chi^2(4) = 116.94, p < .001$
91	5	$\chi^2(6) = 4.56, p > .05$	$\chi^2(4) = 321.46, p < .001$	$\chi^2(4) = 207.68, p < .001$
82	6	$\chi^2(6) = 3.01, p > .05$	$\chi^2(4) = 168.82, p < .001$	$\chi^2(4) = 116.00, p < .001$

Note. Chi squares for rules solved and featured failed to show significance, except for nursery school children. The comparisons between the rules applied and featured and applied and solved all showed significance.

^{a,b}Featured is expected. ^cSolved is expected. ^dComparisons with $df = 6$ are calculated over Rules 2 to 8. ^eComparisons with $df = 4$ are calculated over the Rules 3 and 5 to 8.

is just lagging behind solving and that this may explain the divergence between reasoning in problem solving and generation at a certain grade level. To test this hypothesis, we compared the frequency distribution of rules between the generation condition for a group of a given level with the solving condition of each lower level. The results showed that the frequencies of rules differed significantly for all possible pairs (we did not overlap the CPM and the SPM) in a range from $\chi^2(1-4) = 26.25-195.01, p < .001$. Hence, generating in the older children did not bear a similarity to solving in the younger ones. Furthermore, we compared the frequency distribution between the grade levels within each problem condition. Within the solving condition no significant differences were found between grades. However, within generating all differences between grade levels turned to be significant, results ranged from $\chi^2(1) = 6.77, p < .05$, to $\chi^2(1) = 178.96, p < .001$. These results indicate that generation is a cognitive ability in which large developmental shifts occur between nursery school and Grade 6 (ages of 4 to 12 years, approximately).

Components. Children who performed the CPM generated significantly less nonfigurative components than those who performed the SPM ($W_x = 15.00, n_1 = 5, n_2 = 3, p < .05$, two-tailed). With increasing grade level, the score for components and specifications combined increased, $r_s = .667, n = 8, p < .05$, indicating an increasing complexity in the items created. In what way the contents of these items related to those of the RPM is analyzed in the following.

For children who received the CPM, the distribution of the nine geometric components for *featured* and *solved* did not differ, except for the younger nursery school children, $\chi^2(3) = 28.05, p < .001$ (calculated over *line, triangle, cross, and dot*). Three comparisons between the distributions of *featured* and *applied* were

significant, $\chi^2(4) = 82.10, p < .001$, for the younger nursery school children (*diamond, cross, oval, and dot* excluded), $\chi^2(5) = 20.86, p < .001$, for the older nursery school children (*square, diamond, and oval* excluded), and $\chi^2(6) = 16.831, p < .05$, for Grade 2 (*diamond and oval* excluded). The distributions for *solved* and *applied* showed a difference for all grades, results ranged from $\chi^2(5) = 12.23, p < .05$, for Grade 3, to $\chi^2(5) = 62.11, p < .001$, for the younger nursery school children (*diamond, beam, and oval* excluded from both calculations).

For those children who performed the SPM, all comparisons between the distributions of components for *featured, solved, and applied* failed to show a difference. This means that, unlike for *rules*, a difference between generation and solving for *components* failed to show in the three highest grades.

Specifications. Children who received the CPM showed differences between the distributions of the seven specifications for *featured* and *solved*, except Grade 3, $\chi^2(5) = 19.51, p < .01$ (*orientation* excluded). The differences between *featured* and *applied* had significance levels ranging from $\chi^2(5) = 16.47, p < .01$, for the older nursery school children, to $\chi^2(5) = 30.96, p < .001$, for Grade 2 (*orientation* excluded from both calculations). As expected, differences between *solved* and *applied* were all significant, ranging from $\chi^2(5) = 27.28, p < .001$, for Grade 2, to $\chi^2(5) = 67.74, p < .001$, for the younger nursery school children (*orientation* excluded from both calculations).

Children who received the SPM showed no differences between the distributions of *featured* and *solved*. Between *featured* and *applied* the only significant difference was found in Grade 6, $\chi^2(5) = 11.56, p > .05$ (*speckled* excluded). As expected, the differences between *solved* and *applied* were all significant, ranging from $\chi^2(4) = 14.35, p > .01$, for Grade 5, to $\chi^2(5) = 29.74, p < .001$, for Grade 6 (*speckled* and *orientation*

excluded from the former calculation, *speckled* from the latter).

We may conclude that the frequencies of specifications applied in the CRT do not follow those of the items previously encountered in the solving task. The same seems true for the frequencies of components, with the exception of Grades 4 to 6. This result may be due to the creation of a larger proportion of figurative components by children who performed the CPM as compared to the ones who performed the SPM. Figurative items have no further subdivisions for components, while the categories for non figurative items do not allow for the inclusion of new components.

Finally, we observed that with advancing grade level, generally, more rules, and components and specifications were featured in the items created. This indicates that children in higher grades are capable of handling an increasing complexity of rules and richness of objects.

GENERAL DISCUSSION AND CONCLUSION

Participants of different school grades were asked to complete an RPM test before creating an RPM item themselves in the CRT. We interpreted the RPM scores as their standard problem solving ability and the CRT scores of items created as their creative reasoning ability. In many tasks, for instance category decisions (Ward et al., 1989), younger children have been found more likely to focus on one attribute, and older individuals combine several. Accordingly, in this study, complexity of an RPM item was not defined by type of rule, but by the number of rules that constituted it (cf. Jaarsveld, 1994, 2007).

Correspondence was observed between the increasing complexity of problem solving and problem creation. In the RPM, we observed with advancing grade level an increasing number of solutions in items combining several rules. In the CRT there is a parallel increase in the number of relations applied per item created. Children in more advanced grades also used more components, which answered to an increasingly rich variety of specifications.

Although for both RPM and CRT, scores increased with grade level, they were shown to be unrelated within grades, except in Grades 3 and 6. From these results, we infer that for most grades the scores of the problem solving task are not predictive of performance in the creative reasoning task. The exceptions appear not to be systematic, and this may be a first hint that standard problem solving ability is not a pre-condition for creative reasoning.

A more detailed analysis revealed the profound discrepancy between rules featuring in the RPM items and those applied in the CRT. Although CRT rule

preferences were characteristic of grade level, these preferences were based, neither on items shown, nor on items solved in the RPM. So, even though children have immediately prior achieved a traditional problem solving task, which requires a full understanding of how to manipulate matrix components according to certain rules, they still refrained from using these manipulations in item creation.

The gap between solving and creating cannot be understood as absence of a random subset of rules of the solving task from the creation task; otherwise we still would have observed rule preferences to correlate between tasks. Neither does the gap amount to a lag; otherwise, we would have observed correlations in rules preferred between creations of a given grade level with solutions of earlier grades. Rule preferences characteristic of the creative reasoning task appear to be drawn from episodic and semantic memory resources (Cacciari et al., 1997; Ward, 1995; Weisberg, 1993; Wilkening et al., 1997). A particular instance of this is the preference for idiosyncratic rules in younger children.

The same as for rules seems to apply to the component objects to which the rules apply. Even our youngest participants generated components that had little relation with the type presented to them in the RPM. Despite the nonfigurative character of the RPM items, figurative components were persistently preferred in generation.

Overall, participants preferred to introduce rules and other elements from their individual episodic/semantic knowledge domains, as opposed to what they encountered in the problem solving task. This difference cannot be understood as a discrepancy in knowledge domain, as both tasks used the same knowledge base as their starting point. Thus, we must conclude that the discrepancy between classical and creative problem solving is real and not a by-product of differences between knowledge domains from which tasks are typically drawn. Creative problem solving, therefore, does not depend entirely upon classical problem solving skills.

Is the discrepancy primarily a matter of rules or of components? The prominence of figurative components in the generation task may simply be a by-product of the preference for idiosyncratic rules. On the other hand, it could also be explained independently, on the basis of greater accessibility of figurative material in semantic or episodic memory (Cacciari et al., 1997; Ward, 1995; Weisberg, 1993; Wilkening et al., 1997). Greater accessibility of figurative components has parallel advantages for the traditional solving task. Vodegel-Matzen (1994) constructed parallel items to the SPM ones with figurative components only, such as hats, bananas, or faces. These items did not change their relative position in the difficulty ranking as a result, but became easier to solve. In this study, however, no relation was found

between SPM scores and the generation of nonfigurative components in the CRT. Participants who generated components of a figurative character did not score lower on the solving test. We may therefore conclude that in our study, the preference for figurative items is a consequence of rule preference rather than for ease of their availability.

The generation of nonfigurative components increased somewhat with grade. But their type and specifications revealed considerable discrepancies to those encountered in the solving task. Components created differed especially from those featured and solved in nontexture specifications, which allowed for variations in size, number, and orientation. This is a further indication that pictorial characteristics are made subordinate to operational aspects, the latter being directly related to the application of rules.

The fact that the discrepancy observed is focused on manipulations and rules more than on items and their specifications suggests that the differences could be understood in terms of problem spaces. In these terms, the manipulations are active transformations that may bring a problem closer to a solution; components and their textural specifications are the passive materials on which these manipulations occur.

Rules that may lead to the solution in a well-defined problem space are understood as given: Often they are explicitly defined as the rules of the game, whereas for the RPM they are implicit, and assumed to be specified by the problem. Generation in ill-defined problems involves rules that have to be invented. This may require children to spell out the reasoning that could be used in its solution. In this case, doing and knowing cannot be dissociated from each other (cf. Karmiloff-Smith, 1992). To explain the difference between our solving and the creative reasoning tasks, we may therefore consider a gap between *knowing how* and *knowing that*. On the other hand, it is unlikely that this is the major dividing line between our tasks. Given that the classical task taps formal reasoning skills, we may expect that with increasing grade, children would become more apt at spelling out the rules that lead to solutions in the RPM. In that case, we should have found a systematic increase in correlation with grade. However, no such trend was observed in our data.

Ill-defined problem solving in general, and productive reasoning in particular, are usually assumed to require divergent thinking (Guilford, 1956, 1987). This cognitive ability involves, among others, considering several possibilities and exploring remotely related ideas to produce a yet unknown item (Karmiloff-Smith, 1990; Simonton, 1988). We may speculate that greater emphasis on divergent reasoning abilities is one aspect of the independence of classical and creative reasoning. However, as Runco (2003) argued, a design problem cannot be solved by

divergent activity alone. Besides divergent thinking, ability is needed to bring ill-defined problems to an end. Solutions will be valued only as long as they are functional in the context for which they were intended (Akin & Akin, 1998). The set of created possibilities has to be narrowed, rather than searched through (Greeno & Simon, 1988).

Solving an ill-defined problem, therefore, also involves what Guilford (1956) called convergent thinking (Basadur, Ellspermann, & Evans, 1994; Getzels & Csikszentmihalyi, 1976; Hayes, 1989; Liu, Bligh, & Chakabrati, 2003; Mumford, Baughman, & Sager, 2003; Smilansky, 1984; Smilansky & Halberstadt, 1986; Wertheimer, 1945/1968). Convergent thinking within an indeterminate space involves combining created possibilities into a valuable product (Dorst & Cross, 2001; Gero & McNeil, 1998; Jaarsveld, 2007; Jaarsveld & van Leeuwen, 2005), interpreting and implementing criteria, and evaluating created possibilities (Weber & Perkins, 1989). On the aforementioned grounds, it might be argued that the tasks employed in classical creativity tests, such as, the Remote Association Test (Mednick, 1962) and the Collage Design Task (Amabile, 1979), tacitly involve convergent thinking, whereas in these tests these processes are not measured separately.

Both convergent and divergent thinking processes need to cooperate to arrive at a quality formulation (Jaarsveld, 2007; Jaarsveld & van Leeuwen, 2005). Creative thinking contains divergent and convergent thinking abilities, the former to produce new ideas and the latter to evaluate these ideas (Cropley, 2006). In other words, creative thinking is bipolar, containing knowledge apprehension and knowledge utilization (Basadur, 2005); perhaps this is central to the independence of classical and creative problem solving.

This means that, with divergent abilities as an independent core, constraints from convergent reasoning are added, selectively and incrementally, as creative reasoning skills develop. In accordance with this understanding, we observed convergent activity in the CRT to develop relatively slowly. Each grade was characterized by a particular preference for one specific type of rule. Participants within a grade predominantly came up with the same type of rules for their created items. The rule applied in each grade may give us an indication about creative cognitive development that cannot be inferred from the solving task. The dominance of rule 1, *idiosyncratic and semantic coherence*, shows that creative reasoning in the youngest children is dominated by rules that are not deducible logically and clearly arise from an individual interpretation. It is likely that at this age creative reasoning is relatively unconstrained by logical (Klahr, Fay, & Dunbar, 1993; Wilkening et al., 1997) or convergent thinking (Albert, 1996). Instead, the

dominant convergent reasoning appears to be coherence based on a story or anecdote relevant to the child's biography. This focus shifts for the remaining 5 years of secondary education to a preference for rule 3, *continuous pattern*. Deleting a piece in an overall pattern, whether figurative or nonfigurative, may be the first abstract relation that plays a role in generative problem solving. Neither these preferences, nor their development across grades, could be predicted from the solving task. Hence, our understanding about the cognitive abilities of school children would be biased if it is based only on standard problem solving tasks, as in the RPM. Creative reasoning may develop from a core of divergent operations; but even the convergent operations in creative reasoning differ persistently from those in classical reasoning tasks.

REFERENCES

- Akin, O., & Akin, C. (1998). On the process of creativity in puzzles, inventions, and designs. *Automation in Construction*, 7, 123–138.
- Albert, R. S. (1996). Some reasons why childhood creativity often fails to make it past puberty into the real world. In M. Runco (Ed.), *Creativity from childhood through adulthood: The developmental issues* (pp. 43–56). San Francisco, CA: Jossey-Bass.
- Amabile, T. M. (1979). Effects of external evaluation on artistic creativity. *Journal of Personality and Social Psychology*, 37, 221–233.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Barron, F. (1970). The disposition towards originality. In P. E. Vernon (Ed.), *Creativity* (pp. 273–288). London: Penguin.
- Basadur, M. (2005). Modelling applied creativity as a cognitive process: Theoretical foundations. *Korean Journal of Thinking and Problem Solving*, 15, 13–41.
- Basadur, M., Ellspermann, S., & Evans, G. (1994). A new methodology for formulating ill-structured problems. *Omega International Journal of Management*, 22(6), 627–645.
- Boden, M. A. (1990). *The creative mind: Myths and mechanisms*. London: Abacus.
- Cacciari, C., Levorato, M. C., & Cicogna, P. (1997). Imagination at work: Conceptual and linguistic creativity in children. In T. Ward, S. Smith & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 145–178). Washington, DC: American Psychological Association.
- 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.
- Carey, L. J., & Flower, L. (1989). Foundations for creativity in the writing process: Rhetorical representations of ill-defined problems. In J. A. Glover, R. R. Ronning & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 283–304). New York: Plenum.
- Chan, D., & Fookkee, C. (1994). Theories of conditional reasoning: The interpretative component involved in reasoning. *Cognition*, 53, 217–238.
- Chi, M. T. H., & VanLehn, K. A. (1991). The content of physics self-explanations. *Journal of the Learning Sciences*, 1, 69–105.
- Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal*, 18, 391–404.
- Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem-solution. *Design Studies*, 22, 425–437.
- English, L. D. (1998). Children's reasoning in solving relational problems of deduction. *Thinking and Reasoning*, 4, 249–281.
- Fink, A., Grabner, R. H., Benedek, M., Reishofer, G., Hauswirth, V., Fally, M., et al. (2009). The creative brain: An investigation of brain activity during creative problem solving by means of EEG and fMRI. *Human Brain Mapping*, 30, 734–748.
- Garner, W. R. (1970). The stimulus in information processing. *American Psychologist*, 25(4), 350–358.
- Gero, J. S., & McNeill, T. (1998). An approach to the analysis of design protocols. *Design Studies*, 19, 21–61.
- Getzels, J. W. (1975). Problem finding and the inventiveness of solutions. *Journal of Creative Behavior*, 9(1), 12–18.
- Getzels, J. W., & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. New York: Wiley.
- Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science*, 16, 395–429.
- Greeno, J. G., & Simon, H. A. (1988). Problem solving and learning. In R. C. Atkinson, R. J. Herrnstein & G. Lindsey (Eds.), *Stevens' handbook of experimental psychology. Vol 2. Learning and cognition* (pp. 626–672). New York: Wiley.
- Guilford, J. P. (1956). The structure of intellect model. *Psychological Bulletin*, 53, 267–293.
- Guilford, J. P. (1987). Creativity research: Past, present, and future. In S. Isaksen (Ed.), *Frontiers of creativity research* (pp. 33–65). Buffalo, NY: Bearly.
- Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hayes, J. R. (1989). Writing research: The analysis of a very complex task. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 209–234). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Jaarsveld, S. (1994). *Puzzles solving and inventing*. Master thesis, University of Amsterdam.
- Jaarsveld, S. (2007). *Creative cognition: New perspectives on creative thinking*. Kaiserslautern, Germany: University of Kaiserslautern Press.
- Jaarsveld, S., & van Leeuwen, C. (2005). Sketches from a design process: Creative cognition inferred from intermediate products. *Cognitive Science*, 29, 79–101.
- Jausovec, N., & Jausovec, K. (1997). EEG activity during the performance of complex mental problems. *International Journal of Psychophysiology*, 36(1), 73–88.
- Jausovec, J. (2000). Differences in cognitive processes between gifted, intelligent, creative, and average individuals while solving complex problems: An EEG study. *Intelligence*, 28, 213–237.
- Lachmann, T., & van Leeuwen, C. (2008). Different letter-processing strategies in diagnostic subgroups of developmental dyslexia. *Cognitive Neuropsychology*, 25, 730–744.
- Luo, J., & Knoblich, G. (2007). Studying insight problem solving with neuroscientific methods. *Methods*, 42, 77–86.
- Kapa, E. (2007). Transfer from structured to open-ended problem solving in a computerized metacognitive environment. *Learning and Instruction*, 17, 688–707.
- Karmiloff-Smith, A. (1990). Constraints on representational change: Evidence from children's drawings. *Cognition*, 34, 57–83.
- Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.
- Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, 25, 111–146.
- Liu, Y. C., Bligh, T., & Chakrabarti, A. (2003). 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.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, 69, 220–232.

- Mestre, J. P. (2002). Probing adults' conceptual understanding and transfer of learning via problem posing. *Applied Developmental Psychology*, 23, 9–50.
- Mumford, M. D., Baughman, W. A., & Sager, C. E. (2003). Picking the right material: Cognitive processing skills and their role in creative thought. In M. Runco's (Ed.), *Critical creative processes* (pp. 19–68). Cresskill, NJ: Hampton Press.
- Newell, A., & Simon, H. A. (1972). The theory of human problem solving. In A. Newell & H. Simon (Eds.), *Human problem solving* (pp. 787–868). Englewood Cliffs, NJ: Prentice Hall.
- Pine, K., & Messer, D. (1999). What children do and what children know: Looking beyond success using Karmiloff-Smith's RR framework. *New Ideas in Psychology*, 17, 17–30.
- 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.
- Qiu, J., Li, H., Yang, D., Luo, Y. J., Lie, Y., et al. (2008). The neural basis of insight problem solving: An event-related potential study. *Brain and Cognition*, 68, 100–106.
- Raven, J., Raven, J. C., & Court, J. H. (1998a). *Raven manual: Section 1, general overview*. Oxford, England: Oxford Psychologists Press. (Original work published 1938)
- Raven, J., Raven, J. C., & Court, J. H. (1998b). *Raven manual: Section 2, coloured progressive matrices*. Oxford, England: Oxford Psychologists Press.
- Raven, J., Raven, J. C., & Court, J. H. (1998c). *Raven manual: Section 3, standard progressive matrices*. Oxford, England: Oxford Psychologists Press.
- Raven, J., Raven, J. C., & Court, J. H. (1998d). *Raven manual: Section 4, advanced progressive matrices*. Oxford, England: Oxford Psychologists Press.
- Raven, J., Raven, J. C., & Court, J. H. (1998e). *Raven manual: Section 5*. Oxford, UK: Oxford Psychologists Press.
- Reiter-Palmon, R., Mumford, M. D., Boes, J. O., & Runco, M. A. (1997). Problem construction and creativity: The role of ability, cue consistency, and active processing. *Creativity Research Journal*, 10, 9–23.
- Reiter-Palmon, R., Mumford, M. D., & Threlfall, K. V. (1998). Solving everyday problems creatively: The role of problem construction and personality type. *Creativity Research Journal*, 11, 187–197.
- 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., & 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.
- Runco, M. A. (2007). *Creativity. Theories and themes: Research, development, and practice*. Amsterdam, New York: Elsevier.
- Schraw, G., Dunkle, M. E., & Bendixen, L. D. (1995). Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology*, 9, 523–538.
- Silvia, P. J. (2008a). Another look at creativity and intelligence: Exploring higher-order models and probable confounds. *Personality and Individual Differences*, 44, 1012–1021.
- Silvia, P. J. (2008b). Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan (1965). *Creativity Research Journal*, 20, 34–39.
- Siegel, S., & Castellan, J. (1988). *Nonparametric statistics for the behavioural sciences*. New York: McGraw-Hill.
- Siegler, R. S. (2005). Children's learning. *American Psychologist*, 60, 769–778.
- 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.
- Simonton, D. K. (1988). Creativity, leadership, and chance. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 386–428). New York: Cambridge University Press.
- 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.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of intelligence*. New York: Cambridge University Press.
- Strohschneider, S., & Guss, D. (1998). Planning and problem solving: Differences between Brazilian and German students. *Journal of Cross Cultural Psychology*, 29, 695–716.
- Torrance, E. P. (1962). *Guiding creative talent*. Englewood Cliffs, NJ: Prentice-Hall.
- Urban, K. K., & Jellen, H. G. (1995). *The test of creative thinking—drawing production (Test zum Schöpferischen Denken—Zeichnerisch)*. Lisse, The Netherlands: Swets Test Services.
- Vodegel-Matzen, L. B. L. (1994). *Performance on Raven's progressive matrices*. Ph.D. Thesis, University of Amsterdam.
- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 11, 381–419.
- Ward, T. B. (1995). What's old about new ideas? In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 157–178). Cambridge, MA: MIT Press.
- Ward, T. B., Vela, E., Peery, M. L., Lewis, S. N., Bauer, N. K., & Klint, K. A. (1989). What makes a vibble a vibble? A developmental study of category generalization. *Child Development*, 60, 214–224.
- Weber, R., & Perkins, D. (1989). How to invent artifacts and ideas. *New Ideas in Psychology*, 7(1), 49–72.
- Weisberg, R. W. (1993). Creative problem solving. In R. W. Weisberg (Ed.), *Beyond the myth of genius* (pp. 91–124). New York: Freeman.
- Wertheimer, M. (1968). *Productive thinking (Enlarged edition)*. London: Tavistock. (Original work published 1945)
- Wilkening, F., Schwarzer, G., & Rümmele, A. (1997). The developmental emergence of multiplicative combinations. In T. B. Ward, S. M. Smith & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 111–128). Washington, DC: American Psychological Association.
- Yamamoto, Y., Nakakoji, K., & Takada, S. (2000). Hands-on representations in two-dimensional space for early stages of design. *Knowledge Based Systems*, 13, 357–384.

APPENDIX A: THREE BOOKLETS OF THE RAVEN PROGRESSIVE MATRICES

The RPM is a set of nonverbal intelligence tests that are often applied to measure intelligence independent of language processing, to predict school performance and to measure general cognitive abilities, e.g., in dyslexia (Lachmann & van Leeuwen, 2009). The booklets are constructed using a subset of sets A, Ab, B, C, D, and E. The sets determine the booklet's level of difficulty, as each set is more difficult to solve than the previous one. Moreover, each set consists of 12 items, which get more difficult to solve as one progresses through the series.

Coloured Progressive Matrices test (CPM; Raven, Raven, & Court, 1998b). The CPM test is designed to assess the cognitive abilities of young children, mentally retarded persons, and the elderly (Raven Manual: Section 2, 1998). This test contains 36 items and comprises sets A, Ab, and B. The items within the three sets present six alternative answers below the given matrices. Set A presents a relation, which consists of a continuous pattern, applicable to all components equally. Success in Set A depends on a person's ability to complete continuous patterns that, toward the end of the set, change in one, and then in two, directions at the same time (Raven et al., 1998a). Set Ab has been interpolated between sets A and B. Sets Ab and B present relations, which apply to two or four separate

components. All components in these items have the same form and texture and are equally separated. Success in Set Ab and B depends on a person's ability to see discrete figures, which complete the design. The last few problems in Set B are of the same order of difficulty as the early problems in Sets C, D, and E of the Standard Progressive matrices test.

Standard Progressive Matrices test (SPM, Raven et al., 1998c). The SPM was designed to be used with children, as well as adults whatever their education (Raven Manual: Section 3, 1998). This test contains 60 items and is constructed with sets A, B, C, D, and E. The sets C, D, and E consist of relations between three, or more components and present eight alternatives below the matrices.