



Universiteit
Leiden
The Netherlands

Dynamic testing and excellence: unfolding potential

Vogelaar, B.

Citation

Vogelaar, B. (2017, January 18). *Dynamic testing and excellence: unfolding potential*. Retrieved from <https://hdl.handle.net/1887/45569>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/45569>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/45569> holds various files of this Leiden University dissertation.

Author: Vogelaar, B.

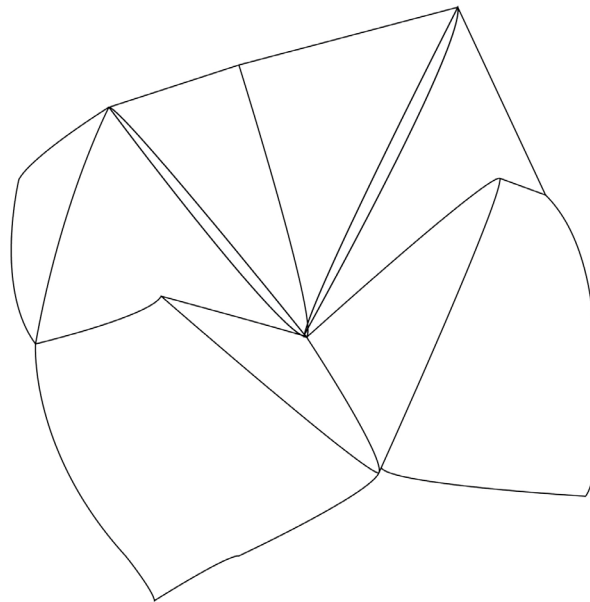
Title: Dynamic testing and excellence: unfolding potential

Issue Date: 2017-01-18

CHAPTER 2

Gifted and average-ability children's progression in analogical reasoning in a dynamic testing setting

2



Bart Vogelaar
Wilma C. M. Resing

To be published as:

Vogelaar, B., & Resing, W. C. M. (2016). Gifted and average-ability children's progression in analogical reasoning in a dynamic testing setting. *Journal of Cognitive Education and Psychology*, 15(3).

Abstract

This study sought to provide more insight into potential differences in progression of analogical reasoning comparing gifted with average-ability children taking into account age, using a dynamic testing approach, using graduated prompting techniques, in combination with microgenetic methods. The participants were between the ages of 5 and 8 years old and were divided into 4 subgroups: gifted unguided control (n = 37), gifted dynamic training (n = 41), average-ability unguided control (n = 95) and average-ability dynamic training (n = 93). We predicted that gifted and average-ability children would show differential progression in analogical reasoning, benefit differentially from a dynamic training procedure, and would show differential instructional needs. The two “ability categories” (i.e., gifted vs. average-ability) were found to show similar, rather than differential, progression paths, and to benefit from a training procedure, whereas gifted children outperform their average-ability peers in accuracy at each session. Likewise, no differences in need for instruction were found amongst these two groups. In general, moreover, younger children seemed to have lower accuracy scores, progress less and need more help than older children. Implications of these findings for the research field of giftedness as well as for education of the gifted and talented are considered in the discussion.

2.1 Introduction

In many educational systems in the world, emphasis has traditionally been on average-ability students. In various countries, it is still seen as controversial, at the very least, that gifted students might require, or benefit from, special educational needs (Persson, 2010). It is often taken for granted that gifted children will somehow manage their classroom learning, and do not need any help or extra attention (De Boer, Minnaert, & Kamphof, 2013). This view, however, seems to be changing; currently, there is more attention for the presumed needs of these children (e.g., Robinson & Olly, 2014), although tailoring the learning to their specific needs is still challenging (Reis & Renzulli, 2010). Given this current interest in the education of gifted children, the main aim of our study was to find out whether, and if so, how children identified as gifted differ from average-ability children regarding their potential for learning and their need for instruction in a classroom setting.

Because of the diverse nature of the body of research into giftedness, and education of gifted children (Dai, Swanson, & Cheng, 2011), research into this field is challenging (VanTassel-Baska, 2006). One of the challenges in the field of research on giftedness is that there is no agreement among researchers on a definition of this concept (Dai & Chen, 2013). What generally seems to be agreed on, however, is that gifted persons have exceptional cognitive capacities (e.g., Renzulli, 2002), and, in addition, a heightened capacity for solving complex problems (Sternberg, 2001). Children are often identified as being gifted by means of conventional, static, and often shortened intelligence tests (Kline, 2001; Lohman & Gambrell, 2012). However, the idea that conventional, static intelligence measures may not always lead to valid and reliable outcomes has been known for some time (Budoff, 1987). Opponents of these tests argue that they predominantly test previously acquired knowledge and skills (Elliott, Grigorenko, & Resing, 2010), which means, for example, that children with a low socio-economic status, a different cultural background, or with special needs can be disadvantaged on these tests (Elliott, 2003; Grigorenko, 2009; Serpell, 2000). Children with a different ethnic background often grow up in different environments, having less preschool education, and different expectations of their parents (Calero et al., 2013; Peña, 2000; Resing, Tunteler, de Jong, & Bosma, 2009; Tzuriel & Kaufman, 1999). As a result, they have less knowledge and skills required for achieving excellent static test scores. A consequence of this is that children's cognitive abilities and intellectual potential may not be accurately portrayed (Elliott, Lidz, & Shaughnessy, 2004).

As a response to the shortcomings of static tests, dynamic testing has been proposed as an alternative of, or supplement to conventional tests (Haywood &

Lidz, 2007; Lidz & Elliott, 2000; Sternberg & Grigorenko, 2009). Dynamic testing is a form of testing that assumes to measure a child's potential for learning (Budoff, 1987; Resing & Elliott, 2011), while incorporating individualized feedback and instruction in the testing process (Elliott, 2003; Jeltova et al., 2007), and measuring a child's improvement after feedback/help has been given. In this way, these tests have the potential to provide in-depth insight into the learning process and development of children (Grigorenko, 2009) as well as the underlying processes involved in learning. Because individualized feedback and instruction are intertwined in the testing process (Elliott, 2003), it is assumed that dynamic testing has the potential to create a more reliable profile of a child's performance level, cognitive strengths, and weaknesses (Jeltova et al., 2007). This individualized approach to instruction and feedback has been assumed to provide a more reliable picture of future academic performance than using static tests only (Elliott et al., 2010). The possibility of measuring the potential for learning or the processes involved in learning new skills make this form of testing a potentially interesting tool for devising educational strategies and interventions (Jeltova et al., 2007).

Whereas a wealth of research has shown the beneficial value of the application of dynamic testing in special populations, such as children with a low socio-economic status or ethnic minorities, and special needs children such as learning disabled, over the past decades, only a few studies have focused on using dynamic tests with regard to giftedness (Boling & Day, 1993; Calero, García-Martin, & Robles, 2011; Passow & Frasier, 1996) and the placement of gifted children into talented programs (Lidz & Macrine, 2001; Matthews & Foster, 2005). The results did suggest that dynamic tests can be used to assess the learning abilities of gifted children. Even more importantly, Kanevsky's (2000) research among preschool children has shown that gifted children have a more extensive zone of proximal development, the ability to learn new skills faster, and are better at generalizing new knowledge obtained. The learning of gifted children was also found to show high levels of motivation, metacognition, self-regulation, and cognitive flexibility (Calero, García-Martín, Jiménez, Kázen, & Araque, 2007). Moreover, Calero et al. (2011) found that gifted children between 6 and 11 years old showed more progression from pre-training to post-training, started at a higher performance level, and showed significantly more improvement than their non-gifted peers. In summary, the studies stated earlier found that gifted children had a higher learning capacity and potential than non-gifted children, and in some cases, children would not have been identified as gifted if static tests only had been used.

Arguably, the largest difference between conventional, static tests and dynamic testing is that, in the former, instruction is often prescribed and part of

the standardized administration process, whereas in the latter the focus is on children's improvement in their performance after explicit training or assistance. In-depth examination of children's responses to these forms of training or assistance is, according to proponents of dynamic testing, of added value to our understanding of the nature of children's learning (Grigorenko, 2009; Jeltova et al., 2011). A form of dynamic testing that specifically enables investigating the need for instruction is the graduated prompts approach (Campione & Brown, 1987; Resing, 2000). These highly structured techniques not only incorporate specific problem-solving skills and strategies but also include training metacognitive skills such as planning and monitoring (Campione, Brown, & Ferrara, 1982). A more recent study has shown that graduated prompts techniques provide additional information about children's potential for learning by comparing the minimum number of prompts, and investigating differences in the number of metacognitive and cognitive prompts children received in solving problems (Resing et al., 2009).

To make dynamic testing even more insightful regarding these processes, dynamic testing procedures could be combined with microgenetic methods of measurement. Microgenetic research methods, developed to examine both spontaneous, unprompted development and changes in children's cognitive abilities (Siegler & Crowley, 1991), include several measurements within a relatively short and sensitive time frame and focus on individual changes in performance on a single cognitive task (Steiner, 2006). Microgenetic methods have, for example, been used successfully in providing more insight into age-related developments, such as the ability to solve analogy problems. Analogical reasoning, a form of inductive reasoning (Barnett & Ceci, 2002), is considered of crucial importance to the acquisition and application of knowledge (Pellegrino & Glaser, 1982), and solving problems (Chi, Glaser, & Rees, 1982). It is generally known that older children are better at solving analogies than younger children (Csapó, 1997; Hosenfeld, van den Boom, & Resing, 1997). Siegler and Svetina's (2002) microgenetic study showed that while although 6-year-old children's initial analogical reasoning ability was found to be lower than their older peers, after unguided practice, these children's analogical reasoning abilities were found to be similar to 7- and 8-year-old children's abilities. Microgenetic studies among gifted children are, however, limited (e.g., Steiner, 2006).

Microgenetic designs, notwithstanding, have the limitation that they cannot provide a full picture of the dynamics involved in change (e.g., Granott & Parziale, 2002; Siegler, 2006). Combining microgenetic techniques with dynamic testing could, therefore, shed more light on the dynamics of change. Only a few studies, nevertheless, have incorporated both unguided practice and a training procedure (e.g., Alexander et al., 1989; Hosenfeld, Van der Maas, & van den

Boom, 1997; Tunteler, Pronk, & Resing, 2008). These studies have revealed that training in addition to unguided practice can have an added value to children's progression in analogical reasoning. In this study, we combined two approaches, unguided practice and dynamic testing, aiming to examine whether two groups of children,¹ a group identified as gifted by their teachers and parents, and a group of average-ability children profited differently from unguided practice, and the intervention provided by dynamic testing, hoping to obtain more insight into differences in performance of visuospatial analogical reasoning tasks and differences in instructional needs.

Our first cluster of research questions concerned changes over time in the progression of accuracy scores when comparing children identified as gifted and average-ability children, taking into account age. First, we focused on the potential effects of unguided practice. Taking into account ability, we expected that the children identified as gifted would outperform their average-ability age-mates in accuracy scores regarding both initial reasoning ability and progression paths. We therefore hypothesized significant differences in pre-test scores between both groups of children; children identified as gifted starting with higher performances on the Pretest 1, a main effect of unguided practice, and an interaction effect of unguided practice, considering the progression from Pretest 1 to Pretest 2, whereby children identified as gifted would show more progression as a result of unguided practice than the average-ability children (Calero et al., 2011). We also hypothesized a main effect of age; younger children were expected to profit less than older ones (Siegler & Svetina, 2002). Then, we focused on the potential effects of dynamic testing. A main effect of treatment was hypothesized, trained children showing more advanced progression paths in accuracy when solving analogies (Resing, 2000; Stevenson, Hickendorff, Resing, Heiser, & de Boeck, 2013). Furthermore, we hypothesized an interaction between treatment and ability; trained children identified as gifted would show more advanced progression compared to the average-ability trained children (Kanevsky, 1990).

Our second cluster of research questions concerned potential differences in instructional needs amongst children identified as gifted and average-ability children, taking into account age. We hypothesized that the children identified as gifted would need less help to solve the analogies than their average-ability age-mates, and that they would, more specifically, need less cognitive help, because general, metacognitive help would, presumably, suffice in order for them to accurately solve the problems given. This hypothesis builds upon Kanevsky's (1994) findings that gifted children were more responsive to feedback and that they were assumed to have an advantage in self-regulation (see also Calero et al., 2007; Zimmerman, 1989). We further explored whether age would play a role in

the instructional needs of all dynamically tested children and expected that considering that older children were found to be better at solving analogies (Csapó, 1997; Hosenfeld, Van den Boom et al., 1997), they would need less help than the younger children, regardless of their ability.

2.2. Method

Participants

Two hundred and sixty-six participants took part in the study, aged 5-8 years old ($M = 6.23$ in years, $SD = 13.40$ in months), ranging from 5 years and 1 month to 8 years and 10 months in age, 128 boys, and 138 girls. Four children were excluded from the analyses because they did not participate at all measurement moments. All participants spoke Dutch, and went to 1 of 12 regular primary schools in various parts of The Netherlands at the time of testing. The age of 5-8 years old was chosen, because previous research has shown that analogical reasoning skills are developed at this age (Tunteler & Resing, 2007). In this study, children were categorized as “identified as gifted” if both their parents and teachers judged their child to be gifted and were all enrolled in gifted, or talented programs. A second group of children were classified as “average-ability”. Seventy-eight children were categorized as identified as “gifted”, 188 as “average-ability”. Written permission was obtained from the schools and the parents prior to participation in the study.

Design

This study used a three-sessions repeated measurements randomized blocking design with two treatment conditions, the Raven. As a measure of initial reasoning ability, three unguided practice sessions, and a short training session (Table 1). Possible differences in initial inductive reasoning ability between conditions were controlled by this randomized blocking procedure based on the children's Raven score, administered before unguided practice Session 1. Blocked pairs of children were randomly allocated to the two treatment conditions (dynamic training vs. unguided control). All sessions took approximately 20-30 mins. Non-trained children participated in the Raven and unguided practice Sessions 1, 2, and 3 but were not provided with the dynamic training procedure.

During the three unguided practice sessions, children were not provided with any feedback. During the training session, however, children were provided with graduated prompts and scaffolds. Children were subdivided into four subgroups: gifted unguided control ($n = 37$), gifted dynamic training ($n = 41$), average-ability unguided control ($n = 95$), and average-ability dynamic training ($n = 93$).

Table 1. Experimental Design; Raven was Administered Before the Test Sessions

Condition	N	Raven	Unguided Practice 1	Unguided Practice 2	Dynamic Training ^a	Unguided Practice 3
Unguided control group	132	X	X	X	-	X
Dynamic training group	134	X	X	X	X	X

Note^a: The children in the dynamic training group received a graduated prompts training session consisting of similar geometric analogies, the unguided control group did not receive a practice, nor a training session.

Materials

Raven. The Raven Progressive Matrices Test (Raven, 1981) was administered to all participants. The raw scores were used as an indication of their fluid intelligence and initial level of analogical reasoning. The Raven test is a non-verbal intelligence test with multiple-choice figural analogies. The Raven test was shown to have a high level of internal consistency, as determined by a Cronbach's alpha of .83 (Raven, 1981). Among the 5- and 6-year-old children, answer sheets were used with pictures of the multiple-choice options from which the children could circle the correct answer to ensure the validity of the collected data scores. The standard testing procedure was used for the 7- and 8-year-old children. Raw Raven scores were used in the analyses instead of standardized scores because no norm scores were available for 5- and 6-year-old children.

Analogy tasks: Tasks and Dynamic Training Procedure. In this study, a series of visuospatial analogy tasks had to be solved, assumed to measure inductive reasoning (Barnett & Ceci, 2002). During the unguided practice sessions, children were provided with series of 20 equivalent, parallel items, composed of geometric analogies of varying difficulty of the type A:B::C:?. All series had different items of comparable item difficulties. The test sessions were equivalent in terms of item difficulty variation, and the order in which the items were presented, but differed in the sense that each test session was composed of new analogy items. These items were a selection of a test battery originally created by Hosenfeld, Van den Boom et al. (1997) and adapted by Tunteler et al. (2008; see Figure 1 for an example). In the construction of all items, six basic geometrical shapes were used: squares, triangles, hexagons, pentagons, circles, and ovals. Each analogy

was constructed by means of five possible transformations: changing position, adding or subtracting an element, changing size, halving, and doubling. The test was administered as a paper-and-pencil test, and the children were asked to draw the correct answers. Because the children were asked to construct and draw their answers themselves, a test session could not have more than 20 items because of time constraints. Test Session 1 (the pre-test) had a high level of internal consistency, as determined by a Cronbach's alpha of .94 for the accuracy scores

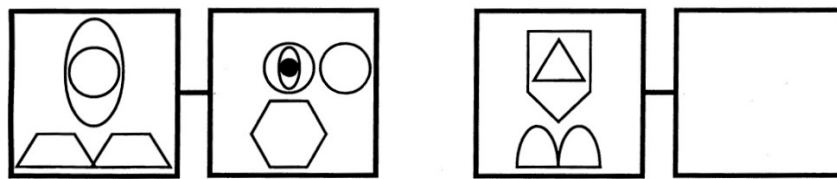


Figure 1. Example of a difficult analogy item.

The training session consisted of 10 geometric analogies. Because the training session took, on average, about 20 mins to conduct, the session could not contain any additional items. None of the items presented in the training session were similar to the items to be solved in the unguided practice sessions. The children in the dynamic training group were given a graduated prompts training (Campione & Brown, 1987; Resing, 2000), a specialized form of dynamic testing consisting of several prompts given to a child when he or she makes an error or a mistake when solving problems. The training procedure was based on Resing's (2000) principles, an adapted form of Campione & Brown's (1987) original graduated prompts approach, and was standardized for all children, containing five steps. Prompts were administered hierarchically: from very general metacognitive prompts to concrete cognitive prompts tailored to the item to be solved. At each step in the solving process, children were asked to draw the solution of the analogy. Each time they drew a solution, they were asked to check their answer. If, after the final step, a child did not succeed in solving the analogy, the test leader provided the child with the correct answer by means of modeling. After having given the correct answer, or having had the correct answer shown by the test leader, for each item, the children were asked to generate a self-explanation: They were asked to explain why they thought their answer was correct. Then, the test leader provided a correct self-explanation, by means of modeling, which included all the transformations necessary to solve the analogy. A schematic overview of the training protocol is provided in Appendix.

General procedure. The children were tested once a week in a period of five consecutive weeks. First, the Raven test was administered in small groups. Then, the unguided practice sessions were administered individually. There were

three unguided practice sessions in total. After the second unguided practice session, the children in the dynamic training group received a short dynamic training session. At the beginning of each test session, the children were given a piece of paper containing the six geometrical shapes used for the analogies. The test leader then named each shape and asked the child to copy the shapes below the printed shapes (analogous to Tunteler et al., 2008). This served three purposes: the children's pre-knowledge regarding the shapes was activated, the test leader and the children both used the same terms for the shapes, and it facilitated the scoring procedure, because the test leader could check which shape the child intended to draw. During the three unguided practice sessions, participants did not receive any feedback on their given answers, nor were they given any help while solving the analogies. The children received minimal instructions only. They were told that they had to solve puzzles with different shapes. Each puzzle had three boxes that were filled and one empty box. The test leader then asked the child which shapes had to be drawn in the fourth box to solve the puzzle.

2.3. Results

Descriptive data

Two one-way analyses of variance with children's initial level of inductive reasoning and age, respectively, as dependent variables and treatment as factor were conducted to evaluate possible differences between children in the two treatment conditions. No significant differences in Raven scores or in mean age were revealed between the two treatment groups, $F(1, 268) = 0.001$, $p = .98$, and $F(1, 268) = 0.45$, $p = .50$, respectively (see Table 2, columns 1 and 2, for mean scores and standard deviations).

Table 2. Mean Scores and Standard Deviations of Raven Scores and Age per Condition

		(1)	(2)	(3)	(4)
		Unguided Control Group	Dynamic Training Group	Gifted Children	Average- Ability Children
	<i>N</i>	132	134	78	188
Raven	<i>M</i>	29.90	29.81	34.00	28.14
	<i>SD</i>	10.97	11.25	9.95	11.11
Age	<i>M</i> in years	6.98	6.85	6.71	6.98
	<i>SD</i> in months	10.56	10.94	9.31	11.41

In addition, two one-way analyses of variance with Raven scores and age were conducted to evaluate initial differences between the two "ability"

categories (gifted vs. average-ability). As expected, the gifted subgroup was found to have significant higher Raven scores than the average-ability subgroup, $F(1, 268) = 16.29, p < .001, \eta_p^2 = .03$. The analysis regarding age revealed no significant differences between the ability categories, $F(1, 268) = 0.36, p = .55$ (see Table 2, columns 3 and 4, for mean scores and standard deviations).

Changes over time in progression of accuracy

Our first cluster of research questions addressed changes over time in the progression of accuracy scores when comparing gifted and average-ability children, and taking into account age. In Table 3, the mean accuracy scores and standard deviations of the children's performance on each of the three unguided practice sessions have been provided, divided by age and subgroup.

Table 3. Mean Scores and Standard Deviations of Analogical Reasoning Accuracy Scores, Divided by Age and Subgroup

		5-6		7-8	
		Gifted Children	Average-Ability Children	Gifted Children	Average-Ability children
		<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Unguided Practice Session 1	Unguided Control Group	8.19 (5.10)	2.42 (3.45)	12.87 (3.87)	9.58 (4.73)
	Dynamic Training Group	6.19 (5.55)	2.44 (3.27)	11.53 (5.93)	10.37 (4.93)
Unguided Practice Session 2	Unguided Control Group	11.77 (6.74)	4.26 (5.84)	17.00 (3.32)	13.37 (5.97)
	Dynamic Training Group	9.81 (7.60)	4.38 (5.37)	14.53 (6.56)	14.11 (5.24)
Unguided Practice Session 3	Unguided control Group	11.18 (7.14)	4.26 (5.72)	17.53 (3.14)	14.60 (5.50)
	Dynamic Training Group	11.77 (7.60)	6.27 (6.84)	17.53 (5.01)	17.06 (3.63)

To examine our first cluster of hypotheses, a repeated measures (RM) analysis of variance (ANOVA) with one within-factor Session (Session 1, Session 2, Session 3) and three between-factors Treatment (unguided control vs. dynamic training), Ability Category (identified as gifted vs. average-ability), and Age (5-6 vs. 7-8 years) was conducted with the number of accurately solved analogy items at the three sessions as the dependent variable. The results showed, as expected,

significant between effects for Ability Category and Age, $F(1, 258) = 33.23, p < .001, \eta_p^2 = .12$ and $F(2, 258) = 107.94, p < .001, \eta_p^2 = .30$, respectively, and a significant main session effect, $F(2, 516) = 151.72, p < .001, \eta_p^2 = .37$. Contrast analysis and visual inspection revealed significant progressions from Session 1 to Session 2 and Session 2 to Session 3, $F(1, 258) = 214.07, p < .001, \eta_p^2 = .45$, and $F(1, 258) = 21.48, p < .001, \eta_p^2 = .08$, respectively, indicating that all groups of children progressed significantly in their accuracy to solve analogies from one session to the next. The RM analysis, as expected, further showed a significant Session x Treatment interaction, $F(2, 516) = 12.62, p < .001, \eta_p^2 = .05$. Contrast analysis showed, as expected, only a significant interaction from Session 2 to Session 3, $F(1, 258) = 21.29, p < .001, \eta_p^2 = .08$, indicating that the dynamically trained groups of children after Session 2 outperformed the groups of children that had unguided practice experiences (see also Figure 2). However, in contrast with our expectations, no significant interactions between Ability Category and Treatment, Ability Category and Session, or Ability Category and Treatment and Session were revealed. These findings indicate that children who were categorized as gifted did have higher scores in analogical reasoning in general but, regardless of whether they were trained or not, showed parallel progression paths when compared with the group of children who were not categorized as gifted. A significant Session x Age effect, $F(2, 516) = 5.25, p = .006, \eta_p^2 = .02$, followed by contrast analysis (only significant from Session 1 to Session 2, $F(1, 258) = 4.37, p < .04, \eta_p^2 = .02$), showed that the age groups only differed in the extent to which they progressed from Session 1 to Session 2, but not from Session 2 to Session 3.

Our conclusion, therefore, has to be that our hypotheses were supported only partially. All groups of children, irrespective of their ability, and age, benefited from both unguided practice and dynamic testing, and, more importantly, dynamic testing led to significantly higher progression in analogical reasoning than unguided practice only. Children's ability category did not mediate these effects.

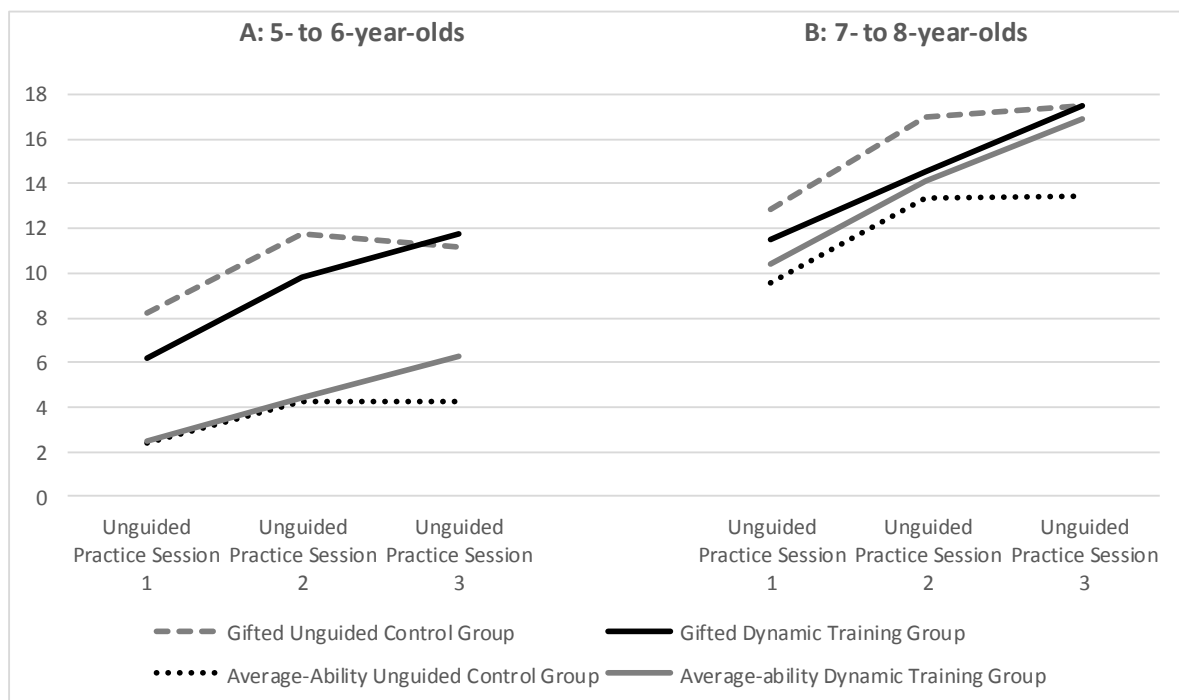


Figure 2. Mean Scores of the Number of Items Correct on session 1, 2, and 3 per subgroup, divided by age in two separate parts for clarity reasons: 5- to 6-year-olds **(A)** and 7- to 8-year-olds **(B)**.

Need for instruction

Our second cluster of research questions addressed differences in instructional needs amongst gifted and average-ability children, while taking into account age. We analyzed the total number of prompts, as well as the metacognitive, and the cognitive prompts the children had received during the dynamic training session. A one-way ANOVA was performed with the total number of prompts as the dependent variable, and Ability Category (gifted vs. average-ability), and Age (5-6 years vs. 7-8 years) as independent variables. As opposed to our expectations, the main effect for Ability Category, $F(1, 127) = 1.13$, $p = .29$, and the Ability Category x Age interaction, $F(1, 127) = .34$, $p = .56$, were not significant, indicating that the children classified as gifted needed approximately similar amounts of help as their average-ability peers, also when taking into account the two age groups. The main effect for Age was, as expected, significant, $F(1, 127) = 17.04$, $p < .001$, $\eta_p^2 = .12$; younger children needed more help in solving analogies (Figure 3).

To research whether the children classified as gifted showed a differential need for metacognitive and cognitive help, a multivariate ANOVA was conducted with the total number of metacognitive, and the total number of cognitive prompts as the dependent variables, on the one hand, and, on the

other hand, Ability Category (gifted versus average-ability), and Age (5-6 years vs. 7-8 years) as independent variables. For both metacognitive and cognitive prompts, the main effect for Age was significant, $F(1,127)=3.98, p<.05, \eta_p^2=.03$, and $F(1, 127) = 26.86, p < .001, \eta_p^2 = .18$, respectively. The analyses, however, did not reveal significant effects for Ability Category or Ability Category x Age effects. These outcomes contradicted our expectations that the children categorized as gifted would need less help, and, in particular, less metacognitive help, irrespective of age. After inspection of the mean scores in Figure 3, our findings led us to conclude that, in general, gifted and average-ability children did not show a differential need for instruction, regarding both the amount and the type of instruction, and that younger children, regardless of ability category, needed more help in general as well as more metacognitive and cognitive help than their older peers.

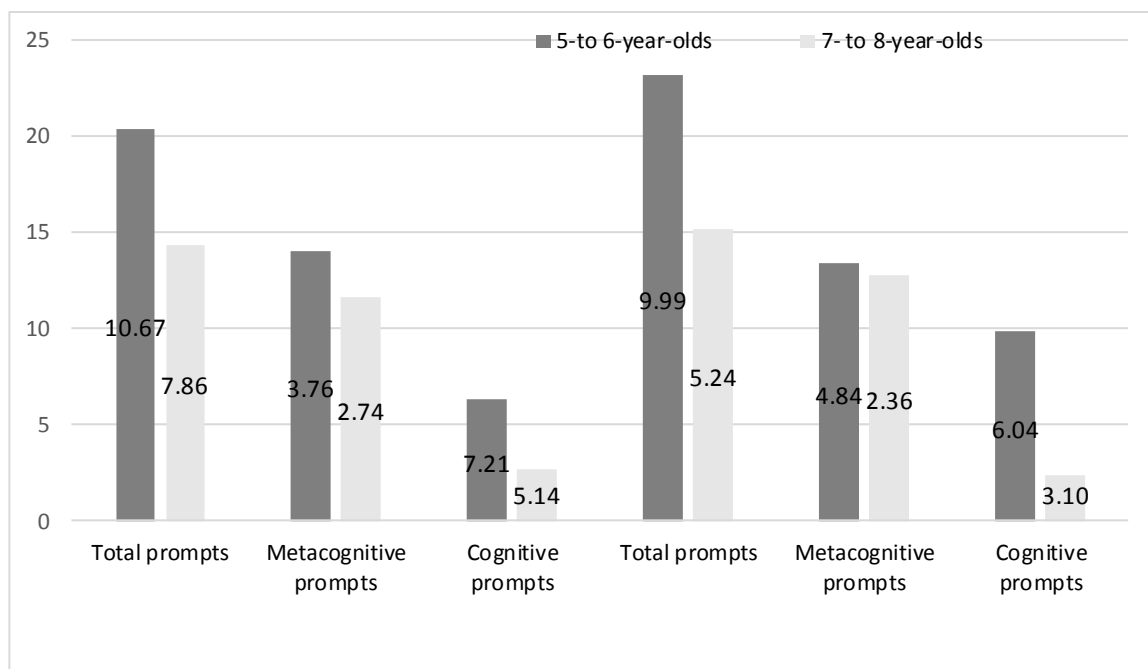


Figure 3. Mean scores and standard deviations (as shown in the individual bars) of the total number of prompts, metacognitive, and cognitive prompts received during training by talent and age.

2.4. Discussion

This study sought to examine whether two groups of children, a group identified as gifted through qualitative judgments by their teachers and parents, and a group of average-ability children, profited differently from unguided practice, and the dynamic testing intervention. Our aim was to obtain more insight into differences in performance of visuospatial analogical reasoning tasks, and differences in instructional needs, taking into account different age groups. Unlike

several studies in the field in which significant differences between gifted and non-gifted children have been found in their performance on a dynamic test (e.g., Calero et al., 2011; Kanevsky, 1990, 1992, 2000; Lidz & Macrine, 2001), and their progression after unguided practice (e.g., Steiner, 2006) were revealed, the children categorized as gifted in this study showed similar, rather than different, progression paths to their average-ability age-mates, while starting at a higher initial ability point than the average-ability and outperforming their average-ability age-mates at each session, regardless of training and age.

Our findings support the idea that microgenetic research methods could lead to additional insight into children's learning, as posited in earlier research (e.g., Siegler, 2006) but that, in this study, they did not show the full picture of change. It seemed that both unguided practice and an additional dynamic training intervention led to progression in children, regardless of ability and age. The progression of the children who did not receive a dynamic training intervention, however, seemed to have stalled after the second session, with no significant increase from the second to the third session. Alexander et al. (1989) found that unprompted performance in geometric analogical reasoning among 4- and 5-year-old children led to a significant increase in performance only after the first session, which, according to these authors, was most probably the result of familiarity with the task. These authors described unprompted geometric analogical reasoning performance of young children as rather stable, which finding seems to be confirmed by this study, even for those children categorized as gifted.

The dynamic testing intervention, however, indeed seemed to lead to additional progression in analogical reasoning from the second session, before training, to the third session, after training, for all groups of children that seemingly could not be explained by practice alone, confirming previous research into graduated prompting techniques (e.g., Resing, 2000; Stevenson et al., 2013). The step-by-step tailored prompts seemed to provide the children with the tools they needed to progress beyond their accuracy scores before training. Looking more closely into the progression from the second session to the third, after training, in line with our expectations, it was found that both the dynamically trained children categorized as gifted, and the children categorized as average-ability benefitted from the dynamic testing intervention. It must be noted, however, that, in contrast with our expectations, and findings from earlier research (e.g., Calero et al., 2011), the children categorized as gifted did not benefit significantly more. Of course, it must be taken into consideration that the group of children, categorized in this study as average-ability, may have included children that, in fact, belonged to the gifted group but were not identified as such, for example,

because of average school results. It is well-documented that children, and in particular gifted children, do not always live up to their potential for excellent performance, potentially as a result of character traits, motivation, internal mediators such as fear of failure, or incorrect usage of strategies (e.g., Reis & McCoach, 2000). It remains as yet unclear, however, whether this might have influenced our research findings.

It must also be taken into consideration that the materials used may not have been sufficiently challenging for the older gifted children, as witnessed by their high mean scores after training. It is possible that there was a moderate ceiling effect, which could also have played a role in the research outcomes regarding potential differences between the ability categories. However, in previous studies using these materials (e.g., Hosenfeld, Van der Maas et al., 1997; Tunteler et al., 2008) children of up to 8 years of age were asked to complete the analogy items. The authors make no mention of a ceiling effect among their older participants, raising the question to what extent this moderate ceiling effect is related to giftedness, to be examined in future research. The latter notion underlines the importance of ensuring that testing and educational material for gifted children is sufficiently difficult (e.g., Kanevsky & Geake, 2004).

Although the older children's results seemed characterized by a ceiling effect, the results of the youngest average-ability children may have been influenced by a bottom effect. In previous studies (Hosenfeld, Van der Maas et al., 1997; Tunteler et al., 2008), the materials used in this study have not been used by children younger than the age of 6 years old. If replicated findings of this study indeed show that among children of 5 years old, there is a bottom effect in accuracy scores, this may mean that important developmental changes at this age are occurring regarding analogical reasoning ability. Of course, at this stage, this is only a speculation that needs to be examined further. In this light, it must be taken into consideration that we employed a short training session only. It would be interesting to conduct future research with a more extensive training procedure and investigate to what extent gifted and average-ability children of different ages would then show differential progression. The fact that both the gifted and average-ability children portrayed similar progression paths can be linked to Steiner's (2006) suggestion that all children's thinking, regardless of ability, develops according to Siegler's (1996) overlapping waves model. This model posits that children of a certain age have access to various strategies to solve problems, and vary in using these strategies over time, while the least effective strategies gradually become disused. In other words, although the gifted children in this study did, in general, outperform their average-ability age-mates, their development was also characterized according to the same principles of varying

strategy choice.

When it comes to the children's performance across the different age groups, in accordance with earlier studies (Csapó, 1997; Hosenfeld, Van der Maas et al., 1997), we found that the younger children's analogical reasoning was characterized by lower initial performance scores, regardless of ability. In addition, our results showed that differential progression paths among the two age groups only occurred from the first to the second session and not from the second to the third, with an advantage for the older children whose progression paths were steeper. It is well-known that great variability exists throughout childhood in the development of children's ability to solve analogies (e.g., Siegler & Svetina, 2002; Tunteler et al., 2008), which becomes apparent through large individual differences within each age group regarding initial ability as well as progression. The fact that the older children showed more progression from Session 1 to Session 2 could be explained, partially, by the fact that through unguided practice in analogical reasoning, children develop various, seemingly more sophisticated, strategies (e.g., Tunteler et al., 2008), and rules, that are more likely to lead to accurate problem solving. Younger children have in previous findings been shown to be more inflexible when it comes to changing to a new strategy or rule, because their ability to execute a new rule or strategy requires inhibiting the old one, and this process is, amongst younger children, still fragile (e.g., Kirkham, Cruess, & Diamond, 2003), which could account for the fact that the progression paths of the older children were steeper from the first to the second session.

Moreover, our findings regarding instructional needs showed that, irrespective of age, the gifted and average-ability children had similar instructional needs, which was in contrast with our expectations and findings from previous research (Calero et al., 2007; Kanevsky, 1990, 1994). Considering that all children who were categorized as gifted in this study attended gifted or talented education, this finding does hold important implications for gifted and talented education. Although it is generally assumed that gifted children manage their own classroom learning (De Boer et al., 2013), because they are said to be self-regulated learners and self-starters (e.g., Azevedo & Hadwin, 2005; Risemberg & Zimmerman, 1992), it seems that this does not necessarily mean that all gifted children have a need for less instruction. This research finding underlines the importance of using instructional and differentiation techniques in gifted and talented education, tailored to individuals' instructional and more general educational needs, for instance, by means of adaptive instruction. This is a type of instruction that aims to increase individual potential through performance demands appropriate for the individual (Heller, 1999). Considering that all the gifted children in this study were enrolled in either talented or gifted education,

a type of education that in general aims to make as much use of high potential as possible (Dai & Chen, 2013), and endeavors to enable independent learning (Heller, 1999), it is surprising that their analogical reasoning progression in this study was not characterized by more independent learning. Future research could investigate this more closely, investigating whether the type of education influences the extent to which a child portrays independent learning, in the hopes of tailoring these types of education even more to the specific needs of talented and gifted children to achieve the best possible fit.

In addition to the limitations mentioned in the preceding text, our study had some other limitations. Our study looked into the learning progression of 5-to-8-year-old children. Because there were no norm scores available for 5-year-old children, we used the raw scores of the Raven instead of percentile or IQ scores. By means of answer sheets with pictures of the multiple-choice options, we safe-guarded the validity of the collected data. Using percentile scores, however, could have been of additional use in categorizing children as gifted or non-gifted, because it might have led to two more distinct groups of children than in this study. As explained earlier, we cannot be entirely certain that our group of average-ability children did not contain any children who did not excel in school but, nonetheless, did have above-average intelligence, in spite of the fact that the two ability categories (gifted vs. average-ability) in our study were found to differ in terms of Raven scores. Of course, in this light, it must be noted that the Raven scores are static, rather than dynamic, scores that have been known to be biased (e.g., Elliott, 2003), and can lead to underestimation of a child's true cognitive abilities (e.g., Jeltova et al., 2007). In future research, categorization into gifted and average-ability groups based on dynamic rather than static measures is advisable. Moreover, if, indeed, moderate ceiling and bottom effects were revealed in our study, one would assume that the group of children experiencing the bottom effect, the 5- and 6-year-old average-ability children would show a need for significantly more instruction, whereas the group of children experiencing the ceiling effect would show they needed significantly less instruction. The reasons as to why the children's instructional needs were not found to differ are as yet unknown and can be investigated further in future studies.

The fact that the gifted children showed progression paths and instructional needs similar to the average-ability children, with variability in progression as well as instructional needs just like the average-ability children, and the fact that gifted children were, in general, found to have higher accuracy scores, ultimately suggests that dynamic testing can be used to measure the potential for learning of all children, including children with higher intelligence. The question that still

needs answering is whether, and if so, to what extent, gifted, talented, and non-gifted children really differ qualitatively regarding their learning characteristics and processes (Dai & Chen, 2013). Our research results underline the importance and usefulness of combining microgenetic research results with dynamic testing procedures and gained more insight into potential differences in analogical reasoning development of young gifted and average-ability children. Hopefully, future research employing microgenetic techniques in combination with dynamic testing procedures could shed more light on this question. Important and promising areas to research in more detail employing these techniques would be strategy use and transfer because these are areas in which gifted children are assumed to differ significantly from non-gifted children in performance (e.g., Kanevsky, 1990; Scruggs & Mastropieri, 1988). Combining these topics in research might lead to a more detailed insight into the learning processes and educational needs of talented and gifted children, which would enhance our understanding of the underlying concepts involved. This, in turn, would greatly inform educational practice of these special groups.

Note

1. Because the children in this study were still very young, between 5 and 8 years of age, the identification of children as gifted was done through qualitative judgments by the children's parents and teachers: a procedure that is often used in order to select children for special talent or gifted educational programs. In The Netherlands, intelligence testing is not standard practice in primary schools, and the identification of young children as gifted is considered controversial.

Appendix. Schematic Overview of the Graduated Prompts Training Protocol