Dynamic Testing of Children's Solving of Analogies: Differences in Potential for Learning of Gifted and Average-Ability Children

Bart Vogelaar
Wilma C. M. Resing
Femke E. Stad
Leiden University, The Netherlands

This study investigated potential differences in the processes of solving analogies between gifted and average-ability children (aged 9–10 years old) in a dynamic testing setting. Utilizing a pre-test-training-post-test control group design, participants were split in four subgroups: gifted dynamic testing ($n = 24$), gifted control ($n = 26$), average-ability dynamic testing ($n = 48$), and average-ability control ($n = 52$). Irrespective of ability group, dynamic testing resulted in a larger number of accurately applied transformations, changes in the proportion of preparation time utilized, and more advanced usage of solution categories. Differences were found between and within the groups of gifted and average-ability children in relation to the different process variables examined.

Keywords: giftedness; graduated prompts; dynamic testing; analogical reasoning

Ⅰn educational settings, when the cognitive abilities of gifted children are to be measured, often (shortened) traditional intelligence tests are used (Pierson, Kilmer, Rothlisberg, & McIntosh, 2012). In such tests, the focus is predominantly on the end products of these tests, in many cases IQ scores. When focusing too much on test scores, however, we run the risk of neglecting the underlying learning and thinking processes that allow children to achieve these scores. It has been argued that analysis of such processes provides useful insights into children's cognitive abilities (Sternberg & Grigorenko, 2002). Indeed, research demonstrates that effective task approaches and solving procedures not only predict performance on cognitive tasks (e.g., Siegler & Svetina, 2002), but also scholastic achievement (Parrila & McQuarrie, 2015; Sternberg & Grigorenko, 2002). Therefore, it may not come as a surprise that high level performance

of gifted children is often ascribed to usage of high level problem-solving steps and procedures (Cho & Ahn, 2003; Muir-Broaddus, 1995).

Conventional (intelligence) tests, however, do not systematically provide information about the processes involved in problem-solving. In addition, since these tests capture previous learning experiences, with children solving tasks independently (Sternberg & Grigorenko, 2002), researchers believe they do not to provide insight into children's learning ability or potential for learning (Elliott, Grigorenko, & Resing, 2010). Previous studies have, moreover, suggested that, irrespective of children's cognitive abilities, their past learning performances or assumed cognitive ability level are not predictive of their potential for learning (Vogelaar, Bakker, Elliott, & Resing, 2017). An alternative means of testing, which taps into children's potential for learning is dynamic testing. In this form of testing, training or extensive instruction are integrated into the testing procedure (Sternberg & Grigorenko, 2002). As dynamic tests utilize skills or abilities that are yet to develop or in development, this form of testing allows for measuring children's learning while testing (Elliott, Resing, & Beckmann, 2018). As such, these tests do not only allow for measuring children's potential for learning, but also for detailed examination of the processes involved in learning (Elliott et al., 2018). The aim of the current study was to investigate whether a dynamic test of analogical reasoning could shed light on differences in the processes of problem-solving of gifted and average-ability children.

DYNAMIC TESTING

Different from traditional, also known as static testing, dynamic testing refers to an interactive approach to psychoeducational assessment, incorporating help, often in the form of feedback or instruction, into the testing procedure (e.g., Sternberg & Grigorenko, 2002). Dynamic testing finds its origin in the ideas of Vygotsky (1978), who posited that children develop and learn within a zone of proximal development (ZPD). According to Vygotsky, learning occurs within the ZPD with children learning with and from other, more capable individuals. The ZPD refers to the distance between the actual level (independent problem-solving, before help has been given) and potential level (problem-solving after help has been given) of development.

As many dynamic tests have a test-training-test design, they are assumed to tap into children's ZPD, providing measures of both the actual (pre-test) and potential (post-test) levels of development, providing insight into an individual child's potential for learning (Elliott et al., 2010; Robinson-Zañartu & Carlson, 2013). These tests often utilize inductive reasoning tasks (e.g., Hessels, Vanderlinden, & Rojas, 2011; Resing, 2013). This form of reasoning is believed to tap into a variety of cognitive and intellectual processes, and play a role in transfer of knowledge, general problem-solving, and everyday learning (e.g., Goswami, 2012; Klauer & Phye, 2008; Richland & Burchinal, 2012; Sternberg & Rifkin, 1979). In the current study, we employed a dynamic test of analogical reasoning, a subtype of the ability to reason inductively. The test items consisted of visuo-spatial geometric analogies of the form A:B::C:D, which consist of a number of geometrical elements (task attributes) that undergo one or several transformations (changes; see Figure 1 in the "Method" section for more information).

Dynamic tests of analogical reasoning with a test-training-test design are sometimes combined with a graduated prompts training (e.g., Campione, Brown, Ferrara, Jones, & Steinberg, 1985; Resing, 2013). This highly structured training procedure involves providing testees with a graduated, hierarchically administered, series of predetermined prompts. Each time a child cannot solve a problem without help from the examiner, a child is provided with a prompt, with each

new prompt becoming more specific. For each training item, prompts range in specificity from metacognitive prompts, geared at metacognitive processes underlying the problem-solving process, to cognitive prompts, tailored to each individual item, to modeling of the correct answer. Its hierarchical structure potentially enables researchers and practitioners to examine children's individual instructional needs (Ferrara, Brown, & Campione, 1986; Sternberg & Grigorenko, 2002).

PROCESS-ORIENTED DYNAMIC TESTING OF INDUCTIVE REASONING

Dynamic tests of inductive reasoning with a graduated prompts training have been used successfully to provide insight into the processes occurring during the solving of inductive reasoning tasks (Resing, 2013; Resing, Xenidou-Dervou, Steijn, & Elliott, 2012). Focusing on process variables enabled researchers in this field to utilize more fine-grained analyses of children's variations in changes in performance than when only focusing on test outcomes, such as accuracy scores. Such process variables include the type and number of transformations (changes) children are able to deal with, verbalizations of the inductive reasoning procedures they used, and also their time-on-task (Resing, Touw, Veerbeek, & Elliott, 2017; Tunteler, Pronk, & Resing, 2008; Tzuriel & Galinka, 2000).

With regard to time-on-task, researchers examined the time children spend on different phases in the process of problem-solving. Investing in the initial phases of the analogical reasoning process is seen as an advanced approach to solving inductive reasoning tasks, which likely leads to success (Klauer & Phye, 2008; Sternberg, 1985). In previous studies, the amount of time children used for preparing to solve an inductive reasoning task (preparation time) relative to the total time it took them to complete the solving of an item (solving time) was used in order to obtain insight into children's reasoning process (Kossowska & Nęcka, 1994; Resing et al., 2012). In the present study, we utilized Kossowska and Nęcka's (1994) approach to analyzing children's reaction times to examine differences between the children in the proportion of time they spent on preparation when solving the analogical reasoning tasks. In addition, researchers have analyzed the different types of solutions children provided by examining, for each individual solution, the proportion of changes of the different task attributes the children could apply accurately (e.g., Resing et al., 2017; Richland, Morrison, & Holyoak, 2006). In studying children's verbal solving of inductive seriation tasks, Resing et al. (2017), for instance, found that all groups of children showed changes in the proportion of transformations they reported accurately, whereby most children showed significant improvements from pre-test to post-test.

Studies regarding differences in the process of solving inductive reasoning tasks of gifted and average-ability children in the analogical reasoning domain remain scarce. Muir-Broaddus (1995), for instance, reported that high-achieving gifted children utilized larger numbers of sophisticated verbalizations than both underachieving gifted and non-gifted age-mates, but were also more likely to use flexibly those verbalizations that were deemed most beneficial for solving novel problems.

DYNAMICALLY TESTING GIFTED CHILDREN'S INDUCTIVE REASONING ABILITIES

A number of studies have shown that dynamic tests are valuable instruments for unveiling the potential of a wide range of children (e.g., Lidz, 1987; Resing, 2013; Tzuriel, 2013). Only few

studies, however, investigated the use of dynamic testing for gifted children (Calero, García-Martín, & Robles, 2011; Kanevsky, 2000), especially with regard to dynamic tests of inductive reasoning. In a study into a dynamic test utilizing analogy items of the form A:B::C:D, four groups of third grade children were studied: gifted, outstanding high performance, outstanding low performance, and typically performing children (Tzuriel, Bengio, & Kashy-Rosenbaum, 2011). Both before and after training gifted children outperformed the others in accuracy, but only differences between the gifted and outstandingly low performing groups of children were significant. In addition, studies by Vogelaar and Resing (2016) and Vogelaar et al. (2017), which both utilized a dynamic test of analogical reasoning, demonstrated that gifted children achieved significantly higher accuracy scores than their average-ability peers at the pre-test, before training, and at the post-test, after training. They also found, in contrast to what they expected, that the extent to which they improved in accuracy was similar.

More importantly, in these three studies it was consistently found that the cognitive abilities of gifted children can be characterized by large individual differences, providing support for the growing notion that gifted children do not form a homogeneous group when it comes to their cognitive capacities (Reis & Renzulli, 2009). Although these studies have demonstrated the value of testing gifted children's inductive reasoning capacities dynamically, no studies have been conducted that systematically investigate the problem-solving processes of gifted children in a dynamic test setting.

THE CURRENT STUDY

The current study aimed to shed light on potential differences between gifted and average-ability children in the processes occurring when solving analogy items in a dynamic test setting. We specifically sought to examine whether a dynamic test with a pre-test-training-post-test format, vis-à-vis a static test consisting of a pre-test and a post-test only, could provide us with additional insight into different aspects in the processes occurring during solving of analogies. Our first research question concerned children's changes from pre-test to post-test in the number of transformations they could apply accurately. We expected (1a) that the children who received dynamic training would show more improvement in accurately applied transformations than their peers in the control group, who received the pre-test and post-test only (Resing & Elliott, 2011; Tunteler et al., 2008). As to the potential differences between gifted and average-ability children, we expected (1b) that both at pre-test and post-test gifted children would show a larger number of accurately applied transformations than their average-ability peers (Calero et al., 2011; Kanevsky, 2000; Steiner, 2006).

Secondly, we examined potential changes in the proportion of preparation time children utilized when solving analogies. It was expected (2a) that all groups of children would spend relatively more time on preparation at the post-test than at the pre-test, with (2b) children who were dynamically trained showing larger changes in the time they spent on preparing their solutions than those in the control group (Resing et al., 2012). Regarding potential differences between gifted and average-ability children, we expected that (2c) the gifted children would show a larger increase from pre-test to post-test in the proportion of time they spent on preparation than their average-ability peers (Cho & Ahn, 2003; Muir-Broaddus, 1995; Steiner, 2006).

Thirdly, the quality of children's solutions provided at pre-test and post-test was examined in terms of the proportion of changes children applied accurately. Similar to previous studies,

children's solutions were divided into three separate categories based on the proportion of transformations applied accurately: "high," "medium," and "low" (e.g. Resing et al., 2017). Based on these research findings, it was expected that (3a) from pre-test to post-test all groups of children would show an increase in the medium and high solution categories, and a decrease in the low solution category. It was further expected (3b) that the children who were trained would show the same changes as their peers in the control group, but that the difference between pre-test and post-test would be more salient for these children. Finally, in relation to differences between gifted and average-ability children, it was expected (3c) that the changes from pre-test to post-test of the gifted children would be more pronounced than their average-ability peers (Muir-Broaddus, 1995).

In addition, we examined whether we could allocate children to specific groups, based on the quality of the solutions they provided (a high, medium or low proportion of accurately applied transformations). We expected (3d) that children would be more likely to be categorized as providing "high" solutions as a consequence of training (Resing et al., 2017). Furthermore, we anticipated that (3e) at the pre-test as well as the post-test gifted children would be more likely to be allocated to the "high" solution group than their average-ability peers (Muir-Broaddus, 1995).

METHOD

Participants

The study employed 150 nine- to 10-year-old participants, *M*age = 9;11, *SD*age = 0;6, 72 boys and 78 girls. The participants were enrolled in primary schools in middle and high SES neighborhoods located in the western part of the Netherlands. Selection of children and schools occurred on the basis of their willingness to participate. All parents had given written informed consent prior to the children participating in the study. In total, 100 children were categorized as average-ability and 50 as gifted. Gifted children were oversampled. Eight children did not complete all sessions of the study due to illness. Their data were removed from the calculations. The overall sample size was calculated by means of a sample size analysis (Hulley, Cummings, Browner, Grady, & Newman, 2013), which revealed a power of 80% to find an effect size of .49 when comparing gifted and average-ability children, and a power of 80% to find an effect size of .46 comparing the two experimental conditions (dynamic testing versus control condition).

Giftedness identification was conducted based on children's scores on the Raven Standard Progressive Matrices (RSPM) (Raven, Raven, & Court, 2000). In accordance with the position statement of the National Association for Gifted Children (2010), children were categorized as gifted if the scored at or above the 90th percentile of the RSPM.

Design

The present study utilized an experimental two-session (pre-test-post-test) control group design with randomized blocking with two conditions: dynamic testing versus control condition (see Table 1 for a schematic overview of the design).

Test sessions were administered once a week, over a total period of five subsequent weeks. All test sessions were administered individually, and took approximately 20–30 minutes each. The children in the dynamic testing condition were given two short training sessions between the pre-test and post-test. The children in the control condition were asked to complete an unrelated control task consisting of paper and pencil dots-to-dot tasks. In these tasks, children were asked

TABLE 1. Overview of the Experimental Design

Condition	Groups	Dynamic versus static test				
		Raven	Pre-test	Training 1	Training 2	Post-test
Dynamic test-ing	Gifted (*n* = 24) Av. ability (*n* = 48)	X	X	X	X	X
Control	Gifted (*n* = 26) Av. ability (*n* = 52)	X	X	CT	CT	X

Note. CT = control task.

to connect the dots on a piece of paper that together form a picture. Administration of the control task took the same amount of time as the training sessions, to ensure that the children in the control condition had a similar amount of one-on-one time with the experimenter as those who received training.

Prior to pre-testing, the RSPM was administered. The RSPM is assumed to provide an indication of children's fluid intelligence, which, in turn, is considered a key component in many conceptions of intellectual abilities (Sternberg, Jarvin, & Grigorenko, 2011). Randomized blocking was utilized per gender and school with children's initial reasoning ability (raw scores on the RSPM), as the independent blocking variable (Kirk, 2013). On the basis of their RSPM scores, per ability group, pairs of children were created and randomly assigned to the two conditions, per school and gender. This procedure was meant to ensure that the two conditions would not differ significantly in inductive reasoning ability. Children were allocated to one of the following subgroups based on our blocking procedure: gifted dynamic testing (*n* = 24), gifted control (*n* = 26), average-ability dynamic testing (*n* = 48), and average-ability control (*n* = 52).

Materials

Raven's Standard Progressive Matrices. The RSPM (Raven et al., 2000) was used as our independent blocking variable. The RSPM is a non-verbal test of inductive reasoning which utilizes visual analogies. The split-half-reliability is *r* = .91 (Raven et al., 2000). In our sample of participants, good reliability was found (Cronbach's α = .88). The 1992 corrected norm scores for the Dutch population were used in the current study to calculate percentile scores.

Dynamic Test of Analogical Reasoning. The dynamic test employed visual-spatial geometric analogies. The items used were originally developed by Hosenfeld, Van den Boom, and Resing (1997), and converted by Tunteler et al. (2008) into separate test sessions. The test, consisting of four sessions in total (a pre-test, two short training sessions followed by a post-test), has a constructed response format. Each item was constructed using the following geometric shapes (the elements): hexagons, pentagons, squares, triangles, ellipses and circles, and the answers had to be drawn by the children. The original items were constructed using a maximum of six different transformations (changes): adding or subtracting an element (doubling), changes in position, changes in size, and halving. As the original item-sets were developed for young children, and
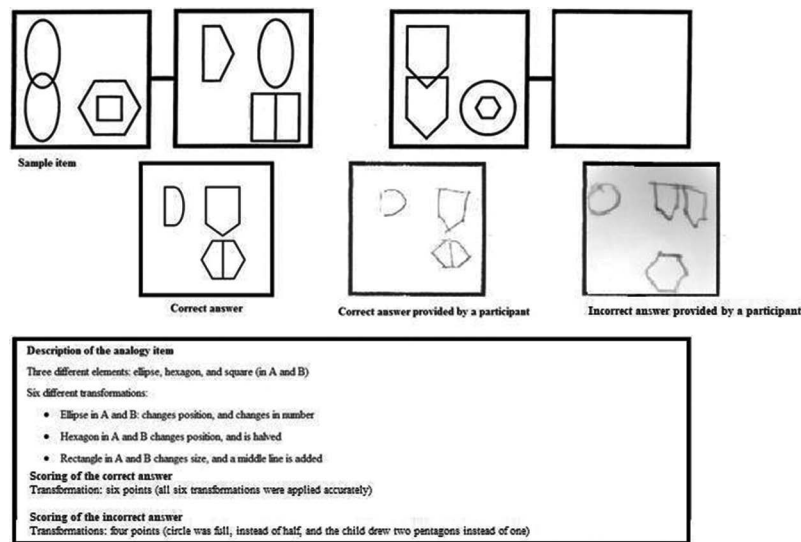
Figure 1. Example of an analogy item, including the correct answer, a correct and an incorrect answer.

the present study employed older children, aged 9–10, the items utilized were made more difficult by adding possible extra transformations: rotation, adding of middle line, and color.

In solving analogies of this type, children have to compare A and B, and identify the transformations (changes in task features). Then, they have to compare A and C, and map the transformations occurring from A to B onto C, in such a way that the correct solution can be constructed in D. A correct solution of such an analogy, therefore, implies that all transformations occurring from box A to B are identified, and applied accurately from box C to D. Item complexity was defined by both the number of elements (the different shapes used for an item) and the number of transformations these elements undergo (changes in the shapes from A to B, and, similarly, from C to D; Mulholland, Pellegrino, & Glaser, 1980; Sternberg, 1977). The items' theoretical level of difficulty was defined on the basis of a simplified equation posited by Mulholland et al. (1980): Difficulty level = 0.5 × Elements + 1 × Transformations.

In Figure 1, a sample item is provided, including the correct solution, a correct and an incorrect solution provided by a participant, and an explanation regarding the different elements and transformations.

Pre-test and Post-test. Both the pre-test and post-test contained 20 items of varying difficulty. The two test sessions were constructed as equivalent parallel versions, as the difficulty level of the items and the sequence of the items were kept equal. At the start of the pre-test and post-test, children received general, short instructions only, telling them to solve the analogy puzzles without any help.

Both test sessions consisted of 20 items, each having a total of 119 transformations that had to be applied accurately. In Appendix A, an overview of the division of transformations, number of elements, and corresponding theoretical difficulty levels is provided. Good internal consistency was found for the pre-test (α = .84). In addition, test-retest reliability analysis revealed a stronger correlation for those in the control condition (r = .87, p < .001) than those in the dynamic testing condition (r = .56, p < .001). Fisher's r to z transformation demonstrated that the two test-retest correlations were significantly different, z = 4.2, p < .001.

Dynamic Training. Children in the dynamic testing condition were provided with a graduated prompts training consisting of two short sessions. Each session consisted of six similar analogy items. The dynamic training utilized a graduated prompts procedure used in several earlier studies (e.g., Resing & Elliott, 2011; Stevenson, Hickendorff, Resing, Heiser, & De Boeck, 2013; Vogelaar & Resing, 2016; Vogelaar et al., 2017). Graduated prompting involved providing children with a series of predetermined, standardized prompts when they made a mistake in solving an analogy item. The training items were similar to those provided in the pre-test and post-test. The provision of prompts occurred hierarchically: starting with four general metacognitive prompts (e.g., "How did you solve the task last time?"), followed by four item-specific cognitive prompts (e.g., "What are the similarities between these two boxes of the puzzle?"). Due to the hierarchy in the provision of prompts, with each prompt becoming more specific, this procedure allowed for measuring the differing degrees of help individual children needed when solving problems.

After provision of a prompt, the children had to draw the answer they believed was accurate, and then check its accuracy. If a child provided an incorrect answer, the experimenter would say that the answer was not correct yet, and that the child would receive help to come to the correct answer. If after a child had been provided with seven prompts, he or she was not able to come to the accurate solution, the eighth cognitive prompt provided was based on modeling. As a final step of the training procedure, the examiner asked the children to explain why they thought the solution they had drawn was correct, after which the examiner modeled an accurate self-explanation. Overall, the prompts provided encouraged children to solve the analogy items analytically; that is, they trained the children to first think about their prior knowledge, then encode the different elements of the analogy, identifying similarities and differences, and find relationships between the different elements of the item.

The graduated prompts utilized were based on task analysis of solving analogies (e.g., Resing, 1993; Sternberg, 1985). The training protocol was highly standardized, written out in full, so all examiners used exactly the same prompts. A schematic overview of the training protocol is provided in Appendix B.

General Procedure

All different sessions of the current study were administered following a protocol, which ensured that all examiners gave the same instructions to the participants. The data were collected by postgraduate psychology students who received extensive training in how to administer and score the dynamic test. Every week during the data collection, the examiners met with the first author to discuss test administration and scoring method, as well as any potential problems arising.

At the start of both the pre- and post-test, the children were given a piece of paper that contained the geometrical shapes that were utilized in the construction of the analogies. They were asked to provide the names of the shapes, and then to draw them underneath the printed ones, trying to copy the original shape as closely as possible. It was assumed that it would not only activate the pre-knowledge of the children, but would also facilitate the scoring procedure, enabling the examiner and the child to use the same words for the shapes.

Scoring

Three aspects of the processes occurring during the solving of analogy items were investigated. We analyzed (a) the number of accurately applied transformations, (b) the proportion of preparation time in relation to the total time spent on solving the analogy items, as well as (c) the classification of children's given solutions based on the proportion of transformations they applied accurately. To estimate the reliability of the coding procedure, the pre-test data were scored independently by two examiners. Inter-rater reliability for the accurately applied transformations in children's drawings was $\kappa = .85$, $p < .001$.

Number of Accurately Applied Transformations. For each given answer on the pre-test and post-test, the number of accurately applied transformations was counted, with a maximum of 119 transformations that the child could have applied accurately in both the pre-test and the post-test.

Proportion of Preparation Time. Solving time per item, registered by means of a stopwatch, started immediately after the item was presented, and ended when the child had drawn an answer and indicated that they had finished. We distinguished between preparation time, the time children used to prepare and think about their solution, and execution time, the time they spent on drawing their answers. Preparation time ended as soon as the child started drawing a shape, which was the start of the execution time variable. The proportion of preparation time was calculated as follows: preparation time divided by the total solving time (preparation time + execution time), multiplied by 100 (Kossowska & NÈ©cka, 1994). High scores were taken to reflect children spending relatively more time on preparation, whereas low scores were taken to reflect children spending relatively less time on preparing their solutions.

Solution Categories. Children's solutions to the analogy items solved during pre-test and post-test were categorized as "high," "medium," or "low" based on the proportion of transformations they had applied accurately. Originally, this system was used to categorize the verbal solutions of series completion tasks, based on the number of transformations mentioned accurately in the children's verbalisations (Resing et al., 2017). In the current study, we adapted this system to the solving (drawing) of visual geometric analogy tasks. Each of the children's solutions on the pre-test and post-test was analyzed with regard to the proportion of transformations the child had applied accurately. Those drawings containing 0%–25% of accurately applied transformations were categorized as "low," those containing 25%–75% as "medium," and those containing 75%–100% of correct transformations as "high."

Solution Subgroups. In addition, all children were assigned to a solution subgroup, based on their solutions (Resing et al., 2017). Five different solution subgroups were distinguished, depending on the percentage of high, medium, or low solutions given. Appendix C shows in more detail the categorization rules for the five different subgroups.

RESULTS

Initial Group Comparisons

Before examining the research questions, potential differences were examined between the two experimental conditions, and the two ability groups. A multivariate analysis of variance (MANOVA) demonstrated that the children in the two conditions (dynamic testing versus control) did not differ significantly in age ($p = .522$) or initial reasoning ability, as measured by their Raven scores ($p = .362$). In addition, it was found that the gifted and average-ability children also

did not differ significantly in age, (p = .436), but, as expected, did show a significant difference in Raven scores, with an advantage for the gifted group (p < .001, ηp^2 = .24). The gifted children had obtained a mean raw score of 46.35 (*SD* = 5.59; *Mpercentile* = 91.60; *SDpercentile* = 2.27), and the average-ability children 39.04 (*SD* = 7.14; *Mpercentile* = 43.16; *SDpercentile* = 25.56).

Differences in the Processes of Solving Analogies

The effect of training on the two process variables was investigated by means of a repeated measures multivariate analysis of variance (RM MANOVA) including Session (Pre-test vs. Post-test) as within-subjects factor, and Condition (dynamic testing versus control) and Ability group (gifted vs. average-ability) as the between-subjects factors. The number of accurately applied transformations, the proportion of preparation time, and the solution categories (low, medium, and high) served as the dependent variables. All multivariate and univariate effects can be found in Table 2.

The multivariate results demonstrated significant Session, Session × Condition, Session × Ability group, and Session × Condition × Ability group effects. Basic statistics for the different variables examined are provided below in Table 3 and Figure 2.

Accurately Applied Transformations. The univariate effects revealed significant Session, and, more interestingly, Session × Condition effects. In accordance with our expectations, these findings indicated, after inspection of the means, that the participating groups of children showed significant improvements in the total number of transformations they had applied accurately, and that dynamically tested children showed more improvement than their peers in the control group. Significant interaction effects of Session × Ability group, and Session × Condition × Ability group were revealed. Contrary to our hypotheses, however, a visual examination of the means indicated that the average-ability children demonstrated more improvement in the number of accurately applied transformations, and benefitted differently from training than their gifted age-mates. The gifted children, however, did apply accurately more transformations than their average-ability peers at the pre- and the post-test, as indicated by a significant between-subjects effect of Ability group, F (1, 146) = 38.55, p < .001, ηp^2 = .21.

Proportion Preparation Time. A significant effect of Session was revealed, indicating that the groups of children, unexpectedly, showed a decrease from pre-test to post-test in the time they spent on preparing their solution relative to the total solving time. The significant interaction effect of Session × Condition indicated that children in the dynamic testing condition showed a more substantial decrease in the proportion of time they spent on preparing their solution than those who were not trained. The non-significant interaction of Session × Ability group further seemed to suggest that the groups of gifted and average-ability children demonstrated similar instead of differential changes in the proportion of time they spent on preparing their solution. However, a significant interaction effect of Session × Condition × Ability group revealed, contrary to our hypotheses, that amongst the untrained children, the gifted children demonstrated a more substantial decrease while for those who were trained, the average-ability children demonstrated a larger decrease in the proportion of time spent on preparation.

Solution Categories. In accordance with our hypotheses, a significant main effect of Session was found for each of the three solution categories. All groups of children showed a decrease in the low and medium solution categories, and an increase in the high solutions. As to the effect of training, a significant interaction effect was found between Session × Condition for the medium and high solutions, indicating differences in changes from pre-test to post-test as a

TABLE 2. Multivariate and Univariate Outcomes of the RM MANOVA for the Different Process Variables: Number of Accurately Applied Transformations, Proportion Preparation Time, and the Solution Categories at Pre- and Post-test

	Wilks' λ	F	p	ηp^2
Multivariate effects				
Session	.31	78.03	< .001	.69
Session × Condition	.85	6.10	< .001	.15
Session × Ability group	.86	5.77	< .001	.14
Session × Condition × Ability group	.92	3.04	.019	.08
Univariate effects				
Transformations				
Session		136.26	< .001	.48
Session × Condition		12.61	< .001	.08
Session × Ability group		18.79	< .001	.11
Session × Condition × Ability group		3.94	.049	.03
Proportion preparation time				
Session		96.62	< .001	.40
Session × Condition		8.67	.004	.06
Session × Ability group		.04	.847	< .001
Session × Condition × Ability group		7.08	.009	.05
Solution categories				
Low solutions				
Session		56.87	< .001	.28
Session × Condition		2.51	.115	.02
Session × Ability group		18.03	< .001	.11
Session × Condition × Ability group		2.80	.097	.02
Medium solutions				
Session		105.81	< .001	.42
Session × Condition		7.90	.006	.05
Session × Ability group		.38	.537	.003
Session × Condition × Ability group		1.16	.284	.01
High solutions				
Session		223.18	< .001	.61
Session × Condition		13.43	< .001	.08
Session × Ability group		17.82	< .001	.11
Session × Condition × Ability group		5.46	.021	.04

result of training for the medium and high solution categories only. The mean scores suggested that children who received training demonstrated a larger increase in the number of solutions categorized as medium and high than those that were untrained.

Finally, we analyzed differences between gifted and average-ability children in relation to the different solution categories, and found significant Session × Ability group interaction effects for those solutions categorized as low and high only. In contrast with our expectations, the mean scores revealed larger changes from pre-test to post-test for the average-ability children than for their gifted peers. Furthermore, a significant Session × Condition × Ability group effect was found for the high solutions. Again, the average-ability children in the dynamic testing condition showed larger changes than their gifted trained peers.

TABLE 3. Basic Statistics for the Number of Accurately Applied Transformations, Proportion Preparation Time, and the Solution Categories at Pre- and Post-test

|		Dynamic Testing		Control	
		Gifted	Average-ability	Gifted	Average-ability
Pre-test					
Transforma-tions	*M (SD)*	91.54 (15.53)	62.73 (28.53)	92.88 (17.30)	63.37 (28.52)
Proportion preparation time	*M (SD)*	32.72 (7.60)	32.79 (13.96)	34.28 (9.19)	31.97 (9.82)
Low solutions	*M (SD)*	.88 (1.15)	4.60 (4.95)	1.19 (1.70)	4.42 (5.17)
Medium solu-tions	*M (SD)*	4.67 (2.78)	6.77 (2.62)	4.15 (2.40)	6.69 (2.83)
High solutions	*M (SD)*	14.46 (3.53)	8.63 (4.68)	14.65 (3.44)	8.88 (4.88)
Post-test					
Transforma-tions	*M (SD)*	105.54 (5.54)	96.60 (17.39)	101.96 (8.52)	79.83 (28.18)
Proportion preparation time	*M (SD)*	25.16 (5.05)	21.61 (6.30)	27.14 (7.68)	29.01 (7.76)
Low solutions	*M (SD)*	.00 (.00)	.50 (2.34)	.27 (.72)	2.10 (4.02)
Medium solu-tions	*M (SD)*	1.79 (1.53)	3.04 (1.44)	2.15 (1.80)	4.92 (2.98)
High solutions	*M (SD)*	18.21 (1.53)	16.46 (2.87)	17.58 (2.00)	12.98 (5.32)
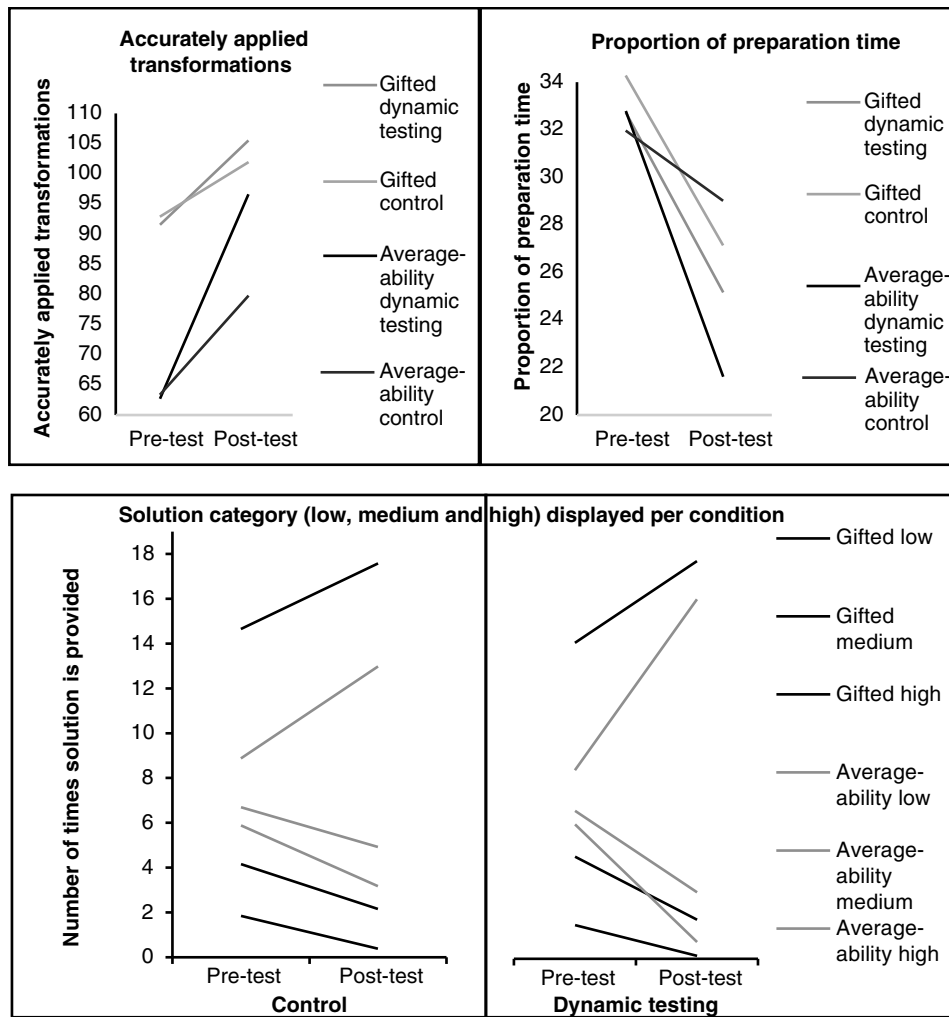
Figure 2. Changes from pre-test to post-test for the number of accurately applied transformations, proportion of preparation time and solution categories divided by condition and ability group.

Exploring Solution Subgroups. The children were allocated to one of five solution subgroups, based on the categorization of their solutions on the pre-test and the post-test. As the solution subgroup variable was measured at an ordinal level, chi-square tests were used to analyze the distribution of the children across the different solution subgroups. The outcomes of a first chi-square test for the pre-test showed, as expected, only significant differences in score distributions for ability groups, with an advantage for the gifted children, $\chi^2(4, N = 150) = 37.70$, $p < .001$, but not for condition, $\chi^2(4, N = 150) = 3.34$, $p = .502$. The outcomes of a second chi-square test for the post-test scores, however, showed a significant difference in score distributions for the children in the two conditions, $\chi^2(4, N = 150) = 19.70$, $p = .001$. The frequencies depicted in Table 4, as hypothesized, suggested that the children who were trained were more likely to be allocated to solution subgroups that provided larger numbers of medium and high solutions than the children who did not receive training.

TABLE 4. Changes in Solution Subgroups from Pre-test to Post-test, Divided by Condition and Ability Group

		Non-inductive	Mixed Partial and Non-inductive	Partial Inductive	Mixed Partial and Full Inductive	Full Inductive	Total
Solution subgroups at the pre-test							
Gifted control	Percentage	0.00	3.85	11.54	34.62	50.00	100.00
Average-ability control	Percentage	23.08	1.92	28.85	34.62	11.54	100.00
Gifted dynamic testing	Percentage	0.00	0.00	8.33	54.17	37.50	100.00
Average-ability dynamic testing	Percentage	29.17	2.08	22.92	41.67	4.17	100.00
Solution subgroups at the post-test							
Gifted control	Percentage	0.00	0.00	0.00	15.38	84.62	100.00
Average-ability control	Percentage	9.62	1.92	15.38	32.69	40.38	100.00
Gifted dynamic testing	Percentage	0.00	0.00	0.00	8.33	91.67	100.00
Average-ability dynamic testing	Percentage	2.08	0.00	0.00	14.58	83.33	100.00

Two separate chi-square tests regarding the post-test were conducted for children in the two conditions. The results revealed that gifted and average-ability children were distributed differentially across the five solution subgroups in the control condition, $\chi^2(4) = 15.08$, $p = .005$, with, as expected, significantly more gifted children than average-ability children who were categorized in the medium and high solution subgroups. In the dynamic testing condition, however, gifted and average-ability children were found to be distributed evenly across the five subgroups, $\chi^2(4) = 1.13$, $p = .569$.

DISCUSSION

The present study sought to examine whether dynamic testing could be used to uncover information about the processes occurring when 9- and 10-year-old gifted and average-ability children solve analogies. In doing so, we focused on changes in the number of transformations children could apply accurately, the proportion of preparation time they used and categorization of children's provided solutions.

Firstly, focusing on changes in the processes occurring during solving of analogies, we found that children who were tested dynamically showed larger improvements in the number of accurately applied transformations after training than their non-trained peers, which is in line with the literature (Elliott et al., 2010; Robinson-Zañartu & Carlson, 2013; Sternberg & Grigorenko, 2002). Interestingly, gifted and average-ability children showed significantly different levels of improvement, with average-ability children improving more, and also profiting more from training. The gifted children, did, however, accurately apply more transformations at the pre- and post-test than their average-ability peers, with a less profound difference between these subgroups at post-test than at the pre-test.

In combination with differences in the test-retest correlations, these findings led us to conclude that testing children dynamically results in additional information about their cognitive potential than testing them statically, as the training procedure seemed to tap into children's ZPD. This could also be an explanation for the different values for the internal consistency of the post-test of the children in the two conditions. Detailed examination of the post-test scores demonstrated that some children in the training condition solved the more difficult items correctly, but some of the easier items incorrectly. These variations seemed larger in the dynamic testing than in the control condition. Secondly, our study focused on the proportion of time children spent on preparing the solving of analogies. In contrast with previous research (e.g., Resing et al., 2012), we concluded that children spent relatively less time on preparation at post-test than at pre-test, with larger changes for, on the one hand, non-trained children compared to their trained peers, and, on the other hand, the average-ability children compared to their gifted peers. The unexpected decrease in the preparation time could be related to familiarity with the task. Moreover, Klauer and Phye (2008) concluded in their study on training inductive reasoning processes that in expert (inductive) reasoning spending both longer and shorter amounts of time on preparing to solve a task can lead to effective task solution.

According to these authors, adequate preparation enables one to solve any kind of inductive reasoning problem, but some people, and experts in particular, may prefer to spend little time on preparation which enables them to come to a global, rapid but accurate solution. In this light, it seems plausible that the children in our study, in particular those considered gifted, became such an expert in solving analogies that they could afford to spend only little time on preparing the solving of analogies. In addition, these variations children showed mirror Siegler's (1996, 2007)

notion that children can choose from a variety of approaches when solving a particular problem, suggesting that the children in the current study tried different approaches when solving the analogy items.

Changes in how children solve analogy items was also visible in a decrease in the number of solutions categorized as low and medium, in combination with an increase in those categorized as high. Those who were trained demonstrated larger changes than those who were not trained, which is in line with the findings of Resing et al. (2017) in verbal solutions of children. These findings provided a further indication of the usefulness of dynamic tests in revealing information about children's potential for learning and change. In contrast with what we expected, changes from pre-test to post-test in the proportions of different solution categories were larger for average-ability than for gifted children.

A similar trend was apparent when allocating children to different solution subgroups. Before training started, the gifted children already provided more medium or high solutions than their non-gifted peers. Similar to Resing et al.'s (2017) study, children who were trained were more likely to be allocated to the subgroups accurately applying larger proportions of transformations than their untrained peers. These findings could be due to the ceiling effect mentioned above, which could have led to the gifted children already providing solutions with a large proportion of accurate transformations at the pre-test (e.g., Muir-Broaddus, 1995; Steiner, 2006). Similar findings were reported by Tzuriel et al. (2011), who concluded that differences between the various subgroups when solving analogies decreased after a dynamic training, and postulate that this is due to an equalizing effect of dynamic testing. This effect refers to the potential of dynamic testing to diminish differences in performance of different groups of children. Generally, post-test scores of children from different groups are more equal than their pre-test scores.

An alternative explanation, however, for the findings in dynamic test outcomes related to differences between the gifted and average-ability children lies in the nature of the giftedness categorization procedure used in the current study. Children were categorized as gifted based on score on the RSPM. Usage of the RSPM or other non-verbal intelligence tests in giftedness identification procedures is seen as a robust measure of intelligence, and is advocated in several studies, as these tests have been shown to be less biased toward linguistic and cultural diverse children than verbal tests (Ford, Grantham, & Whiting, 2008; VanTassel-Baska, Feng, & Evans, 2007). Including full scale (dynamic) giftedness identification would perhaps allow for larger differences between the two groups of children.

In the light of the theoretical framework offered by Vygotsky, perhaps some of the children who were categorized as average-ability children were, in fact, high-ability children who would not have been selected as such, based on their static Raven scores. Although we found that the two groups differed significantly on their RSPM scores, it cannot be discounted that some of the average-ability children might not have shown their full potential on the RSPM, and needed a dynamic training procedure to unveil their capabilities.

Secondly, we cannot discount that some of the gifted children encountered a ceiling effect, which might have been related to the average-ability children showing more improvements in accurately applied transformations. Ceiling effects have been found frequently in other dynamic tests used for gifted children in various domains (see Kanevsky, 2000 for an overview). Inspecting the data at a fine-grained level showed that the analogy items used in the current study did not optimally differ in difficulty level for, at least part of the gifted children, in spite of the fact that the items we adapted were rather difficult, containing up to five different elements and 14

transformations. There is some evidence, in addition, that not all transformations are equally difficult for children to process (Siegler & Svetina, 2002), although Mulholland et al.'s (1980) operationalization of item complexity was used successfully in a number of studies to predict accuracy of solving visual-spatial analogies (e.g., Hosenfeld et al., 1997; Stevenson et al., 2013). Differences in how children process different type of transformations, especially with regard to high-ability children, could be the focus of future studies. For such studies, piloting new analogy items with gifted children prior to conducting the study would be highly advisable to prevent a ceiling effect occurring. Another focus of such studies could be enhancing the difficulty level of training items, for example, by tailoring training difficulty to children's ability level.

Such studies might, in addition, investigate the use of digital dynamic testing. This would make the testing procedure less labor and time-intensive, and would allow for more adaptive forms of training, on the basis of testee's real-time quantitative feedback. It would also enable measuring changes in the processes of analogical reasoning in even more detail. For example with regard to the proportion preparation time, Thibaut and French (2016) found that children who are more skilled in analogical reasoning focus their attention more on the A:B of the analogy then on the C:D terms. Digital dynamic testing in combination with eye-tracking, as conducted in the study of Hessels et al. (2011), would enable measuring changes in the focus of attention. If this would be combined with a questionnaire asking participants what activities they under-took while preparing their solution, this would provide additional insight into changes in the proportion of time children use to prepare their solution.

The findings of the current study imply that dynamic testing can be used effectively to obtain insight into the processes of problem-solving of both average-ability and gifted children. Moreover, the current study demonstrated that children's processes when solving analogies are characterized by individual differences, irrespective of their initial reasoning abilities and their assumed ability level. Moreover, our findings suggest that children need training to help them unveil their potential, and that high ability children do not always manage learning on their own, but need additional help to show their capabilities (De Boer, Minnaert, & Kamphof, 2013). Dynamic testing can be used successfully for high-ability children to uncover information about problem-solving processes, and identify their cognitive capacities (Calero et al., 2011). As researchers voice their concerns regarding a disproportionally low number of children from dis-advantaged groups who are identified as gifted as a result of traditional intelligence tests (Cao, Jung, & Lee, 2017), dynamic testing might be a more equitable method of identifying potential.

REFERENCES

Lidz, C. S. (Ed.). (1987). *Dynamic assessment: An interactional approach to evaluating learning potential*. New York, NY: Guilford Press.

Sternberg, R. J., Jarvin, L., & Grigorenko, E. L (Eds.). (2011). *Explorations in giftedness*. New York, NY: Cambridge University Press.

Calero, M. D., García-Martín, M. B., & Robles, M. A. (2011). Learning potential in high IQ children: The contribution of dynamic assessment to the identification of gifted children. *Learning and Individual Differences*, *21*, 176–181. https://doi.org/10.1016/j.lindif.2010.11.025

Campione, J. C., Brown, A. L., Ferrara, R. A., Jones, R. S., & Steinberg, E. (1985). Breakdowns in flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence*, *9*, 297–315. https://doi.org/10.1016/0160-2896(85)90017-0

Cao, T. H., Jung, J. Y., & Lee, J. (2017). Assessment in gifted education: A review of the literature from 2005 to 2016. *Journal of Advanced Academics*, *28*, 163–203. https://doi.org/10.1177/1932202X17714572

Cho, S., & Ahn, D. (2003). Strategy acquisition and maintenance of gifted and nongifted young children. *Exceptional Children*, *69*, 497–505. https://doi.org/10.1177/001440290306900407

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

De Boer, C., Minnaert, A. E. M. G., & Kamphof, G. (2013). Gifted education in The Netherlands. *Journal for the Education of the Gifted*, *36*, 133–150. doi:https://doi.org/10.1177/0162353212471622

Elliott, J. G., Grigorenko, E. L., & Resing, W. C. M. (2010). Dynamic assessment: The need for a dynamic approach. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (Vol. 3, pp. 220–225). Oxford, England: Elsevier.

Elliott, J. G., Resing, W. C. M., & Beckmann, J. F. (2018). Dynamic assessment: A case of unfulfilled potential? *Educational Review*, *70*, 7–17. https://doi.org/10.1080/00131911.2018.1396806

Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning rules: Studies of proximal development. *Child Development*, *57*, 1087–1099. https://doi.org/10.2307/1130433

Ford, D. Y., Grantham, T. C., & Whiting, G. W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Exceptional Children*, *74*, 289–306. https://doi.org/10.1177/001440290807400302

Goswami, U. C. (2012). Analogical reasoning by young children. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 225–228). New York, NY: Springer.

Hessels, M. G. P., Vanderlinden, K., & Rojas, H. (2011). Training effects in dynamic assessment: A pilot study of eye movement as indicator of problem solving behaviour before and after training. *Educational and Child Psychology*, *28*, 101–113.

Hosenfeld, B., Van den Boom, C, D., & Resing, W. C. M. (1997). Constructing geometric analogies for the longitudinal testing of elementary children. *Journal of Educational Measurement*, *34*, 367–372. https://doi.org/10.1111/j.1745-3984.1997.tb00524.x

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., & Newman, T. B. (2013). *Designing clinical research: An epidemiologic approach*. Philadelphia, PA: Lippincott Williams & Wilkins.

Kanevsky, L. S. (2000). Dynamic assessment of gifted students. In K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 283–296). Oxford, England: Elsevier.

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences*. Thousand Oaks, CA: Sage.

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

Kossowska, M., & Nęcka, E. (1994). Do it your own way: Cognitive strategies, intelligence, and personality. *Personality and Individual Differences*, *16*, 33–46. https://doi.org/10.1016/0191-8869(94)90108-2

Muir-Broaddus, J. E. (1995). Gifted underachievers: Insights from the characteristics of strategic functioning associated with giftedness and achievement. *Learning and Individual Differences*, *7*, 189–206. https://doi.org/10.1016/1041-6080(95)90010-1

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, *12*, 252–284. https://doi.org/10.1016/0010-0285(80)90011-0

National Association for Gifted Children. (2010). *Redefining giftedness for a new century: Shifting the paradigm*. Retrieved from www.nagc.org

Parrila, R. K., & McQuarrie, L. M. (2015). Cognitive processes and academic achievement: Multiple systems model of reading. In T. C. Papadopoulos, R. K. Parrila, & J. R. Kirby (Eds.), *Cognition, intelligence, and achievement: A tribute to J. P. Das* (pp. 79–100). London, England: Elsevier Academic Press.

Pierson, E. E., Kilmer, L. M., Rothlisberg, B. A., & McIntosh, D. E. (2012). Use of brief intelligence tests in the identification of giftedness. *Journal of Psychoeducational Assessment*, *30*, 10–24. https://doi.org/10.1177/0734282911428193

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for raven's progressive matrices and vocabulary scales. Section 3: The standard progressive matrices*. Oxford, England: Oxford Psychologists Press.

Reis, S. M., & Renzulli, J. S. (2009). Myth 1: The gifted and talented constitute one single homogeneous group and giftedness is a way of being that stays in the person over timer and experiences. *Gifted Child Quarterly*, *53*, 233–235. https://doi.org/10.1177/0016986209346824

Resing, W. C. M. (1993). Measuring inductive reasoning skills: The construction of a learning potential test. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), *Learning potential assessment: Theoretical, methodological and practical issues* (pp. 219–241). Amsterdam, The Netherlands: Swets & Zeitlinger Inc.

Resing, W. C. M. (2013). Dynamic testing and individualized instruction: Helpful in cognitive education? *Journal of Cognitive Education and Psychology*, *12*, 81–95. https://doi.org/10.1891/1945-8959.12.1.81

Resing, W. C. M., & Elliott, J. G. (2011). Dynamic testing with tangible electronics: Measuring children's change in strategy use with a series completion task. *British Journal of Educational Psychology*, *81*, 579–605. https://doi.org/10.1348/2044-8279.002006

Resing, W. C. M., Touw, K. W. J., Veerbeek, J., & Elliott, J. G. (2017). Progress in the inductive strategy-use of children from different ethnic backgrounds: A study employing dynamic testing. *Educational Psychology*, *37*, 173–191. https://doi.org/10.1080/01443410.2016.1164300

Resing, W. C. M., Xenidou-Dervou, I., Steijn, W. M. P., & Elliott, J. G. (2012). A "picture" of children's potential for learning: Looking into strategy changes and working memory by dynamic testing. *Learning and Individual Differences*, *22*, 144–150. https://doi.org/10.1016/j.lindif.2011.11.002

Richland, L. E., & Burchinal, M. R. (2012). Early executive function predicts reasoning development. *Psychological Science*, *24*, 87–92. https://doi.org/10.1177/0956797612450883

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *94*, 249–273. https://doi.org/10.1016/j.jecp.2006.02.002

Robinson-Zañartu, C., & Carlson, J. (2013). Dynamic assessment. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology* (pp. 149–167). Washington, DC: American Psychological Association. (Vol. 3, pp.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York, NY: Oxford University Press.

Siegler, R. S. (2007). Cognitive variability. *Developmental Science*, *10*, 104–109. https://doi.org/10.1111/j.1467-7687.2007.00571.x

Siegler, R. S., & Svetina, M. (2002). A microgenetic/cross-sectional study of matrix completion: Comparing short-term and long-term change. *Child Development*, *73*, 793–809. https://doi.org/10.1111/1467-8624.00439

Steiner, H.H. (2006). A microgenetic analysis of strategic variability in gifted and average-ability children. *Gifted Child Quarterly, 50,* 62–74. doi:10.1177/001698620605000107

Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, *84*, 353–378. https://doi.org/10.1037/0033-295X.84.4.353

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing*. New York, NY: Cambridge University Press.

Sternberg, R. J., & Rifkin, B. (1979). The development of analogical reasoning processes. *Journal of Experimental Child Psychology*, *27*, 195–232. https://doi.org/10.1016/0022-0965(79)90044-4

Stevenson, C. E., Hickendorff, M., Resing, W. C. M., Heiser, W. J., & De Boeck, P. A. L. (2013). Explanatory item response modeling of children's change on a dynamic test of analogical reasoning. *Intelligence*, *41*, 157–168. doi:https://doi.org/10.1016/j.intell.2013.01.003

Thibaut, J. P., & French, R. M. (2016). Analogical reasoning, control and executive functions: A developmental investigation with eye-tracking. *Cognitive Development*, *38*, 10–26. https://doi.org/10.1016/j.cogdev.2015.12.002

Tunteler, E., Pronk, C. M. E., & Resing, W. C. M. (2008). Inter- and intra-individual variability in the process of change in the use of analogical strategies to solve geometric tasks in children: A microgenetic analysis. *Learning and Individual Differences*, *18*, 44–60. https://doi.org/10.1016/j.lindif.2007.07.007

Tzuriel, D. (2013). Dynamic assessment of learning potential. In M. Mo Ching Mok (Ed.), *Self-directed learning oriented assessments in the Asia-Pacific* (pp. 235–255). Dordrecht, The Netherlands: Springer.

Tzuriel, D., Bengio, E., & Kashy-Rosenbaum, G. (2011). Cognitive modifiability, emotional–motivational factors, and behavioral characteristics among gifted versus nongifted children. *Journal of Cognitive Education and Psychology*, *10*, 253–279. https://doi.org/10.1891/1945-8959.10.3.253

Tzuriel, D., & Galinka, E. (2000). *The conceptual and perceptual analogical modifiability (CCPAM) test: Closed analogies-instruction manual*. Ramat-Gan, Israel: School of Education, Bar-Ilan University.

VanTassel-Baska, J., Feng, A. X., & Evans, B. L. (2007). Patterns of identification and performance among gifted students identified through performance tasks. *Gifted Child Quarterly*, *51*, 218–231. https://doi.org/10.1177/0016986207302717

Vogelaar, B., Bakker, M., Elliott, J. G., & Resing, W. C. M. (2017). Dynamic testing and test anxiety amongst gifted and average-ability children. *British Journal of Educational Psychology*, *87*, 75–89. https://doi.org/10.1111/bjep.12136

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*, 349–367. https://doi.org/10.1891/1945-8959.15.3.349

Vygotsky, L. S. (1978). (Interaction between learning and development In M. Cole, J. Scribner, V. John-Steiner, & E. Souberman (Eds.),. *Mind in society: The development of higher psychological processes* (pp.79-91). Cambridge, MA: Harvard University Press.

Correspondence regarding this article should be directed to Dr. B. Vogelaar, Faculty of Social Sciences, Department of Psychology, Section Developmental and Educational Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: b.vogelaar@fsw.leidenuniv.nl

APPENDIX A

Pre-test and Post-test Construction in Terms of the Number of Elements and Transformations per Item Number

	Item Number								
	1–3	4–8	9	10	11–12	13	14–17	18–19	20
Number of elements	2	2	2	3	3	3	3	4	5
Number of transformations	2	3	4	5	6	7	8	12	14
Difficulty level	3	4	5	6.5	7.5	8.5	9.5	14	16.5

APPENDIX B

Schematic Overview of the Graduated Prompts Procedure

	Type of Prompt	Content of Prompt
1	Metacognitive	Activating task-related prior knowledge + check correct answer
2	Metacognitive	Activating prior knowledge regarding problem-solving strategy + check correct answer
3	Metacognitive	Encoding of A, B, C + check correct answer
4	Metacognitive	Self-regulated initiation strategy + check correct answer
5	Cognitive/task specific	Seeing similarities and differences A, B, C + check correct answer
6	Cognitive/task specific	Finding the relationship between A and B + check correct answer
7	Cognitive/task specific	Finding the relationship between A and C + check correct answer
8	Cognitive/modeling	Step-by-step modeling of correct solution

APPENDIX C

Description of the Solution Subgroups and Categorisation Rules

	Strategy Group	Categorisation Rules
1	Low	Solutions categorized as "low" provided in at least 50% of the items
2	Mixed 1 and 3: mixed low and medium	Both "low" and "medium" solutions provided in at least 50% of the items
3	Medium	"Medium" solutions provided in at least 50% of the items
4	Mixed 3 and 5: mixed medium and high	Both "medium" and "high" solutions provided in at least 50% of the items
5	High	"High" solutions provided in at least 75% of the items