This is the authors'post-print of

Vogelaar, B., Resing, W.C.M., Stad, F.E., & Sweijen, S. (2019). Is planning related to dynamic testing outcomes? Investigating the potential for learning of gifted and average-ability children. *Acta Psychologica*, *196*, 87-95. https://doi.org/10.1016/j.actpsy.2019.04.004

Is planning related to dynamic testing outcomes? Investigating the potential for learning of gifted and average-ability children

1. Introduction

When a child’s cognitive abilities are assessed, for example for giftedness identification procedures, often conventional assessment instruments such as IQ or school aptitude tests are used. The focus of such instruments is usually on a ‘product’: the test scores, and much less on the underlying cognitive and metacognitive processes, for example how a child plans and subsequently structures their solving of the task (Robinson-Zañartu & Carlson, 2013). Various research, however, pointed to the usefulness of such information in deciding what a child needs in terms of instructions for future learning in the classroom and other didactic interventions (Elliott, Resing, & Beckmann, 2018).

Dynamic testing, a form of assessment integrating instruction into the testing procedure, focuses on how much a child can learn within a short time-frame, providing information about a child’s potential for learning and instructional needs, the amount and type of instructions a child requires to demonstrate learning (Robinson-Zañartu & Carlson, 2013). In doing so, these tests provide insight into the underlying processes involved in learning, and also allow for examining the role that cognitive factors, such as a child’s ability to plan, play in children’s learning. In the present study, a dynamic test of geometric analogical reasoning was utilized to examine to what extent dynamic testing can be used to provide insight into the potential for learning of gifted and average-ability children, and analyse the extent to which the children’s planning ability would be related to performance on the dynamic geometric analogical reasoning task, and to instructional needs during dynamic training.

1.1. Dynamic Testing

Many dynamic tests have a pre-test-training-post-test format, allowing for structured measuring of children’s learning progression (Elliott et al., 2018). Dynamic tests with such designs are sometimes combined with a graduated prompts training approach, a technique in which children are provided with a prompt if they show a significant difficulty in solving a task independently (Campione & Brown, 1987; Resing & Elliott, 2011). As the prompts are administered in a hierarchical fashion, becoming more specific each time a new prompt is provided, this approach measures the different degrees of help children need when solving tasks (Resing & Elliott, 2011). Often, dynamic tests utilize inductive reasoning tasks (e.g., Passig, Tzuriel, & Eshel-Kedmi, 2016; Resing & Elliott, 2011), because this form of reasoning is assumed to play a central role in a variety of cognitive skills and processes, such as general intelligence, problem-solving and everyday learning (Richland, Morrison, & Holyoak, 2006). A dynamic test of visual-spatial geometric analogical reasoning, a subtype of inductive reasoning, was used in the current study.

To date, dynamic testing studies for gifted children are sparse. In recent dynamic testing literature, it is postulated that gifted children's learning can be characterized by higher scores, both at the pre and post-intervention level (Calero, García-Martín, & Robles, 2011; Vogelaar, Bakker, Hoogeveen, & Resing, 2017), indicating that their overall performance is significantly better compared to their average-ability peers. The current study aimed to further examine the possibilities of dynamic testing procedures for differentiating between gifted and average-ability children's potential for learning by examining their level of progression in accuracy, the extent to which they show changes from pre-test to post-test measures regarding the number of accurately solved analogy items, and their need for instructions during dynamic training.

1.2. Planning

Studies have shown that executive functions can predict dynamic testing outcomes, such as performance after training (Ropovik, 2014), change from the pre to the post-training stage (Swanson, 2006) and need for instructions during training (Stad, Wiedl, Vogelaar, Bakker, & Resing, 2018). Executive functions are often considered to be a conglomerate of inter-related complex, general cognitive processes that enable control of thinking and behavior (Monette, Bigras, & Guay, 2011). Most of the studies in this field examined the role of working memory (e.g., Swanson, 2006; Stevenson, Heiser, & Resing, 2013), and, to a lesser extent, cognitive flexibility (Stad et al., 2018), and metacognition (Vogelaar et al., 2017). In general, it seems that, by providing children with feedback and help during training, dynamic testing facilitates children with weaker executive functions in showing their cognitive potential. Providing these children with help that, step by step, guides them through the solving process seems to compensate for the role executive functions play in solving inductive reasoning tasks.

However, the role of planning, a higher-order executive function, has received little attention in the dynamic testing literature. Kar, Dash, Das, and Carlson (1993) found that planning skills of fifth-grade children can be improved as a result of a dynamic testing procedure, with most improvement shown by poor planners. Finally, a study in which a dynamic version of the WCST was administered to psychiatric patients with schizophrenia revealed that planning showed stronger relationships with post-test than pre-test measures (Wiedl, Schöttke, Green, & Nuechterlein, 2004). With regard to dynamic testing of inductive reasoning, Cormier, Carlson, and Das (1990) concluded that planning in fourth-grade children was related to their progression in analogical reasoning accuracy, with poor planners showing more improvement from pre-test to post-test than good planners.

Planning is often described as devising a mental representation of a certain problem, choosing a suitable solving strategy, and evaluating to what extent the steps taken were effective (Gligorovic & Buha, 2016), and is suggested to be related to various other core and higher-order executive functions (e.g., McCormack & Atance, 2011). Previous research has suggested that planning can predict scholastic achievement (Bull, Espy, & Wiebe, 2008;

Sesma, Mahone, Levine, Eason, & Cutting, 2009), as well as performance in and development of analogical reasoning (Richland & Burchinal, 2013; Zook, Davalos, DeLosh, & Davis, 2004). It seems that in both adults (D'Antuono et al., 2017) and children (Richland & Burchinal, 2013), those who are good planners subsequently perform well in analogical reasoning tasks, suggesting that planning plays an important role in our capacity to reason inductively. Vice versa, inductive reasoning has also been found to predict planning (Unterrainer et al., 2004, Zook et al., 2004), providing support for earlier research suggesting that executive functioning and fluid intelligence are interrelated (Duncan, Burgess, & Emslie, 1995). With the exception of the study by Cormier et al. (1990), research into the relationship between planning and inductive reasoning has focused on static measures of analogical reasoning. Therefore, the current study sought to examine whether planning is differentially related to static and dynamic measures of analogical reasoning. In addition, because previous research has suggested that high-ability children's planning abilities are better developed than those of average-ability (Zook et al., 2004), the current study aimed to explore whether the hypothesized relationship between planning and dynamic measures of analogical reasoning would be equivalent for gifted and average-ability children.

1.3. The Current Study

The current study had two main aims. First, the study sought to further investigate the usefulness of dynamic testing in obtaining insight into the cognitive potential and instructional needs of gifted and average-ability children. Second, the potential association between planning, dynamic measures of analogical reasoning and instructional needs was examined, investigating whether this relationship would be different for gifted and average-ability children.

To examine the effectiveness of the dynamic intervention, the present study first analysed children's progression in analogical reasoning. We expected that the children who were dynamically tested would show significantly more advanced progression in accurately solved analogies than their peers who completed the pre-test and post-test only (Resing & Elliott, 2011). Furthermore, we expected the gifted children to obtain higher scores at the pre-test and post-test than their average-ability peers, and show more progression in accuracy (Calero et al., 2011).

The second research question focused on children's need for instructions, by examining their need for prompts in total, as well as their need for metacognitive and cognitive prompts separately. A decrease in the required number of total, metacognitive, and cognitive prompts from the first to the second training session was expected, as a result of learning in the first training session (Resing & Elliott, 2011). The gifted children were expected to show a more significant decrease, considering that in previous studies they were shown to be more responsive to feedback than their average-ability peers (Kanevsky, 1994). In general, in line with previous dynamic testing research, gifted children were expected to require fewer prompts as compared to average-ability children (Kanevsky, 1994).

The third research question focused on the unique contribution of planning and ability group, gifted and average-ability, to accuracy of analogical reasoning before and after training. It was expected that children's planning ability would predict accuracy, with those with stronger planning abilities outperforming those whose planning abilities were weaker (Zook et al., 2004). It was expected, however, that after training planning abilities would no longer have such a relationship with their accuracy scores as a result of the hypothesized facilitating effect of the training procedure (e.g., Stad et al., 2018; Stevenson et al., 2013; Vogelaar et al., 2017). Furthermore, it was explored whether this compensating effect of the graduated prompts training would be equally present for gifted and average-ability children. We further examined whether the children's ability group would hold predictive effects on pre-test accuracy and post-test accuracy, above and beyond the effect of planning.

Our final research question focused on the unique contribution of planning and ability group in predicting children's instructional needs. Based on earlier research in which executive functions were found to significantly predict children's need for instructions (Stad et al., 2018), it was expected that planning abilities would predict the number of prompts children needed during training, with children with weaker planning abilities requiring more prompts than their peers with stronger planning abilities. The potential unique predictive effects of ability group on instructional needs were explored.

2. Method

2.1. Participants

In the current study, 148 children between the ages of nine and ten participated, 70 boys and 78 girls, M_{age} = 9.99, SD_{age} = .54. All children attended middle class primary schools in the western part of the Netherlands. Children were selected upon their willingness to participate, and their parents had provided written informed consent prior to their participation. Two children were excluded from the data analysis as they did not participate in each test session.

Two groups of children participated in the current study: a group of average-ability ($n = 98$) and a group of gifted children ($n = 50$). Children were identified as gifted on the basis of parents' as well as teachers' judgments combined with their enrolment in educational programmes for gifted and talented children and a cut-off score at the 90th percentile of the Raven Standard Progressive Matrices (RSPM, Raven, Raven, & Court, 2000). Approval of the university's board of ethics in psychology was sought prior to conducting the study.

2.2. Design

The current study employed a two-session repeated measures randomized blocking design with two conditions: dynamic testing versus control. The RSPM (Raven et al., 2000) was administered as a blocking instrument, before administration of the Tower of London

(ToL) and the pre-test, to ensure that the two experimental conditions were as equal as possible with regard to their initial reasoning abilities. Per ability group, school and gender, pairs of children with equal Raven scores (blocks) were assigned randomly to the two experimental conditions, resulting in four subgroups: gifted dynamic testing (n = 24), gifted control (n = 26), average-ability dynamic testing (n = 48), average-ability control (n = 50).

In the first session of the experiment, all children completed the RSPM, after which, in the second session, they completed the ToL. Then, in the third session all children completed the pre-test. Only the children in the dynamic testing condition received a dynamic training, consisting of two sessions, session 4 and session 5, taking 20-30 minutes each, between the pre-test and post-test. The children in the control condition were asked to complete an unrelated control task, completing dots-to-dots completion tasks and mazes, taking approximately the same amount of time in order to keep the time-on-task exposure as equal as possible. In the final session, session 6, all children completed the post-test.

2.3. Materials

2.3.1. Raven Standard Progressive Matrices. The RSPM (Raven et al., 2000) was administered as a blocking instrument, prior to the pre-test and the ToL. The Standard Progressive Matrices is a non-verbal test of inductive reasoning utilising visual analogy items. A split-half-coefficient of r = .91 has been reported (Raven et al., 2000).

2.3.2. Tower of London. The ToL was used as a measure of children’s planning abilities. The digital Psychology Experiment Building Language (PEBL) version was used in the current study (Mueller & Piper, 2014). The ToL involves rearranging a set of colored discs placed on three rods to match a desired goal, in as few steps as possible. Phillips et al.’s (1999) Trial A was used in the current study, which included eight test items increasing in difficulty as the test progressed. The total number of steps children needed in order to solve the eight items was utilized as a measure of planning, where lower numbers of steps correspond with higher planning abilities. In a study by Piper et al. (2015) test-retest reliability analysis revealed a reliability of $r = .15$ for the total number of steps. Although this reliability may cause some concern about the score distribution, according to the authors, however, an effect size of $d = .08$ indicated consistent reporting in their sample of participants.

2.3.3. Dynamic test of analogical reasoning. The dynamic test utilized in the current study consisted of visual-spatial geometric analogies of the type A:B::C:?. The analogy items were developed by Hosenfeld, Van den Boom, and Resing (1997), and adapted by Tunteler, Pronk, and Resing (2008) for further use. Each analogy item was constructed of a maximum of six different geometric shapes: ellipses, circles, triangles, squares, pentagons and hexagons. The items consisted of between two and fourteen different transformations, including the following eight different transformations: adding or subtracting an element, changing position, changing size, halving, doubling, changing color, and rotation. See Figure 1 for an example item. Children were asked to draw their answers in the empty square.

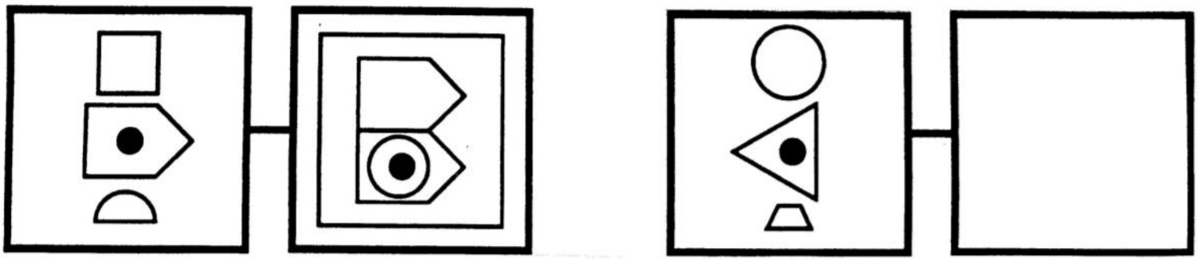

Figure 1. Example of an Analogy Item

2.3.3.1. *Pre-test and post-test.* The pre-test and post-test both consisted of 20 analogy items of varying difficulty, and were constructed as parallel versions with different but equivalent items in terms of difficulty and the order in which they were presented. Prior to solving the analogies at each test session, the children were told to solve the analogy items independently. The pre-test was found to have an internal consistency of $\alpha = .84$, the post-test $\alpha = .95$ for the control group, and $\alpha = .96$ for the dynamic testing group in our sample of participants. Moreover, test-retest reliability analysis revealed a stronger association between pre-test and post-test accuracy scores for the children in the control condition ($r = .70$, $p < .001$), than for those who were tested dynamically ($r = .58$, $p < .001$), providing a preliminary indication of the validity of the dynamic test.

2.3.3.2. Training. The training procedure utilized in the current study employed graduated prompts techniques, as described in earlier studies (e.g., Vogelaar et al., 2017). Training consisted of two sessions of six analogies each. Prompts were provided hierarchically: starting with four general metacognitive prompts, followed by four cognitive prompts tailored to the solution of each individual item. Each new prompt provided was more specific than the previous one, enabling measuring of the differing degrees of help needed by individual children in order to solve the items (Resing & Elliott, 2011). When the child could not provide the correct answer after the seventh prompt, the eighth and final prompt was provided, which consisted of step-by-step guidance to the correct solution (modelling). After solving the item correctly or having been provided with the final prompt, children were asked to explain why they thought the solution provided was correct, followed by the examiner modelling a correct self-explanation.

2.4. Procedure

Children were tested once a week, over a period of six consecutive weeks, taking into account the schools' availability. All different test sessions were administered individually with standardized and protocolled instruction. At the start of the administration of the pre-test and the post-test, children were provided with a sheet of paper containing the six geometric shapes used in the analogy items, and they were asked to first name the shapes and then copy the original underneath the printed shape, staying as close as possible to the original. It was assumed that this procedure would activate the children's prior knowledge, and ensure examiners and participants used the same terminology, facilitating scoring of children's answers (Tunteler et al., 2008).

2.5. Data analysis

To inspect children's progression in accuracy of analogical reasoning and in instructional needs, repeated measures ANOVA was used. We also conducted hierarchical regression analysis to investigate the potential effects on pre-test accuracy, post-test accuracy, and the number of prompts needed during both training sessions. These analyses included the main and interaction effects of condition, ability, and planning, which were all centred to prevent multicollinearity among the independent variables.

3. Results

3.1. Initial Group Comparisons

One-way analyses of variance were performed to inspect possible differences between the two experimental conditions. The children in the control and dynamic training groups did not differ in their age (p = .859), Raven accuracy (p = .502), pre-test accuracy (p = .314) or planning ability (p = .979). No significant differences were found in the distribution of gender across both experimental conditions, $\chi(1)$ = .097, p = .755 and ability groups, $\chi(1)$ = 3.470, p = .062. We also examined possible differences between gifted and average-ability children. No significant differences were found in their age (p = .409) or planning abilities (ToL) (p = .094). However, gifted children, as expected, outperformed the average-ability children in Raven accuracy ($p < .001$, η_p^2 = .234) and in pre-test accuracy ($p < .001$, η_p^2 = .275). Descriptive statistics of these measures per condition and ability group are displayed in Table 1.

		Dynamic training			Control group		
		Gifted	Average-ability	Total	Gifted	Average-ability	Total
	N	24	48	72	26	50	76
Age	*M*	9.94	10.01	9.99	10.15	9.93	10.01
	(*SD*)	(.47)	(.63)	(.58)	(.50)	(.52)	(.52)
Raven accuracy	*M*	47.58	39.83	42.42	46.35	39.14	41.61
	(*SD*)	(5.04)	(6.59)	(7.10)	(5.59)	(7.27)	(7.54)
Pre-test accuracy	*M*	10.04	5.90	7.28	11.31	6.24	7.97
	(*SD*)	(3.89)	(3.45)	(4.08)	(3.06)	(3.79)	(4.28)
ToL	*M*	69.92	71.92	71.25	69.15	72.40	71.29
	(*SD*)	(8.57)	(10.04)	(9.56)	(8.36)	(8.65)	(8.63)

Table 1. *Descriptive Statistics for Age, Raven accuracy, Pre-test accuracy, and ToL Performance, per Condition and Ability Group*

3.2. Effectiveness of Training: Progression from Pre-test to Post-test

To examine children's progression in accuracy of analogical reasoning, a repeated measures ANOVA was performed with Condition (dynamic training/control) and Ability group (gifted/average-ability) as between-subjects factors and Session (pre-test/post-test) as within-subjects factor. A significant main effect of Session indicated that the overall accuracy of all groups of children improved significantly from pre-test to post-test, Wilks' λ = .40, $F(1, 144) = 213.75$, $p < .001$, $\eta_p^2 = .60$. Moreover, a significant Session x Condition effect in combination with a visual check of the mean scores suggested that children in the dynamic training group progressed significantly more in accuracy than children in the control group, Wilks' λ = .89, $F(1, 144) = 17.12$, $p < .001$, $\eta_p^2 = .10$.

However, no significant differences were found between gifted and average-ability children in the steepness of their progression lines from pre-test to post-test, as revealed by a non-significant interaction effect of Session x Ability group, Wilks' $\lambda = .98$, $F(1, 144) = 2.63$, $p = .107$, $\eta_p^2 = .02$. Similarly, a non-significant Session x Condition x Ability group interaction effect, Wilks' $\lambda = .98$, $F(1, 144) = 2.39$, $p = .124$, $\eta_p^2 = .02$, revealed that gifted and average-ability children showed parallel rather than differential progression lines after training or practice at pre-test and post-test only. A significant between-subjects effect of Ability group, $F(1, 144) = 55.37$, $p < .001$, $\eta_p^2 = .28$, revealed that gifted children, as expected, had significantly higher pre-test and post-test scores than their average-ability peers. Figure 2 shows the lines of progression in accuracy of the children in the four subgroups.

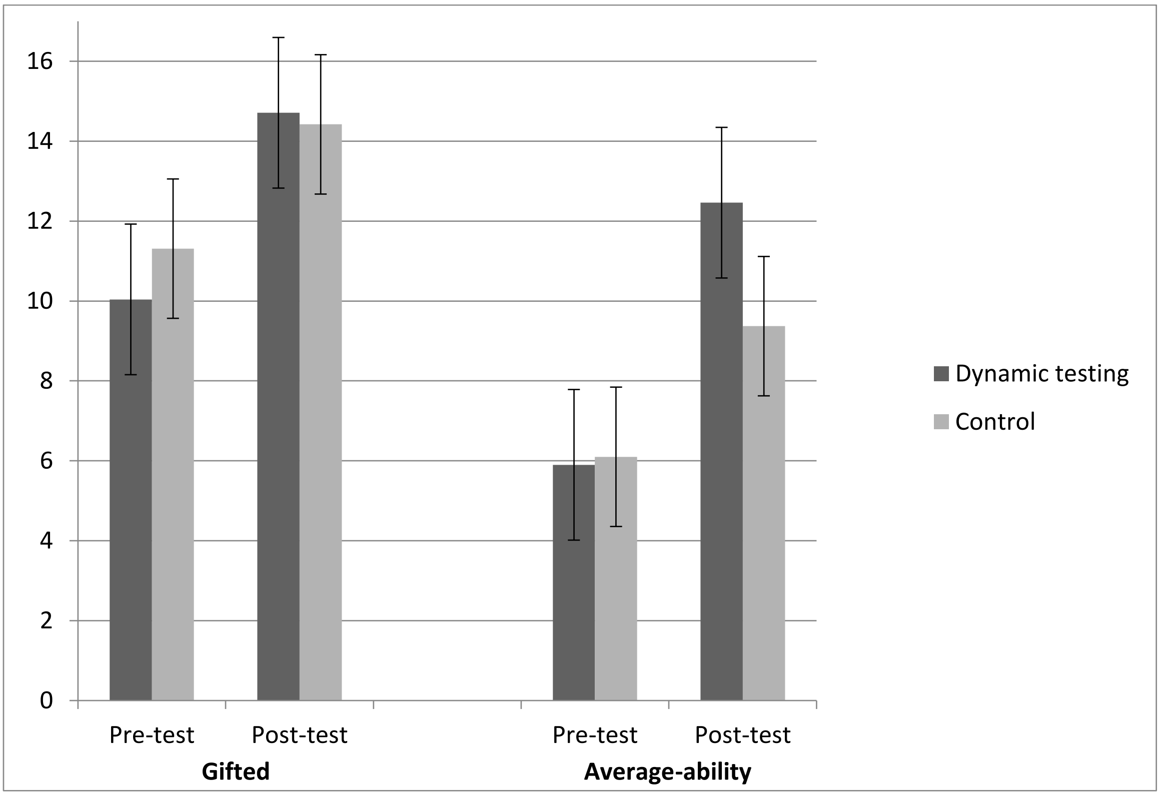

Figure 2. Progression in Accuracy Scores on Pre-test and Post-test per Ability Group and Condition. Error bars denote standard errors.

3.3. Instructional Needs

We also investigated potential differences in children's need for instruction from the first to the second training, and examined the number in prompts in total, as well as metacognitive and cognitive prompts separately. Three separate repeated measures ANOVAs were conducted with Training session (Training 1 – Training 2) as the within-subjects factor, and Ability group (gifted versus average-ability) as between-subjects factor. The number of prompts in total, and the number of metacognitive and cognitive prompts were used, respectively, as the dependent variables. The main and interaction effects of the analyses can be found in Table 2.

	Wilks' λ	*F*	η_p^2
Total prompts			
Session	.86	10.52**	.13
Session x Ability group	1.00	.02	<.001
Metacognitive prompts			
Session	.86	10.85**	.14
Session x Ability group	1.00	.02	<.001
Cognitive prompts			
Session	.94	4.37*	.06
Session x Ability group	1.00	.02	<.001

Note. * $p < .05$, ** $p < .01$

Table 2. *Results of the Repeated Measures ANOVAs for Changes in Children's Need for Instruction from the First to the Second Training Session*

The results revealed a significant main effect of Session for the number of prompts in total, as well as children's need for metacognitive and cognitive prompts. Visual examination of the mean scores, as provided in Figure 3, indicated, in line with our hypothesis, that both gifted and average-ability children showed a decrease from training 1 to training 2 in the number of prompts they had been given in total, as well as for the number of metacognitive and cognitive prompts. However, all three Session x Ability group interaction effects were found to be non-significant. These findings indicated that the gifted and average-ability children showed a similar decrease in their need for instruction from the first to the second training session. However, between-subjects effects of ability group for the total number of prompts, $F(1, 69) = 5.56$, $p = .021$, $\eta_p^2 = .08$, metacognitive prompts, $F(1, 69) = 7.03$, $p = .010$, $\eta_p^2 = .09$, and cognitive prompts, $F(1, 69) = 2.60$, $p = .112$, $\eta_p^2 = .04$, suggested, in combination with visual examination of the mean scores, that gifted children needed fewer prompts in total, and fewer metacognitive prompts, but a comparable number of cognitive prompts.

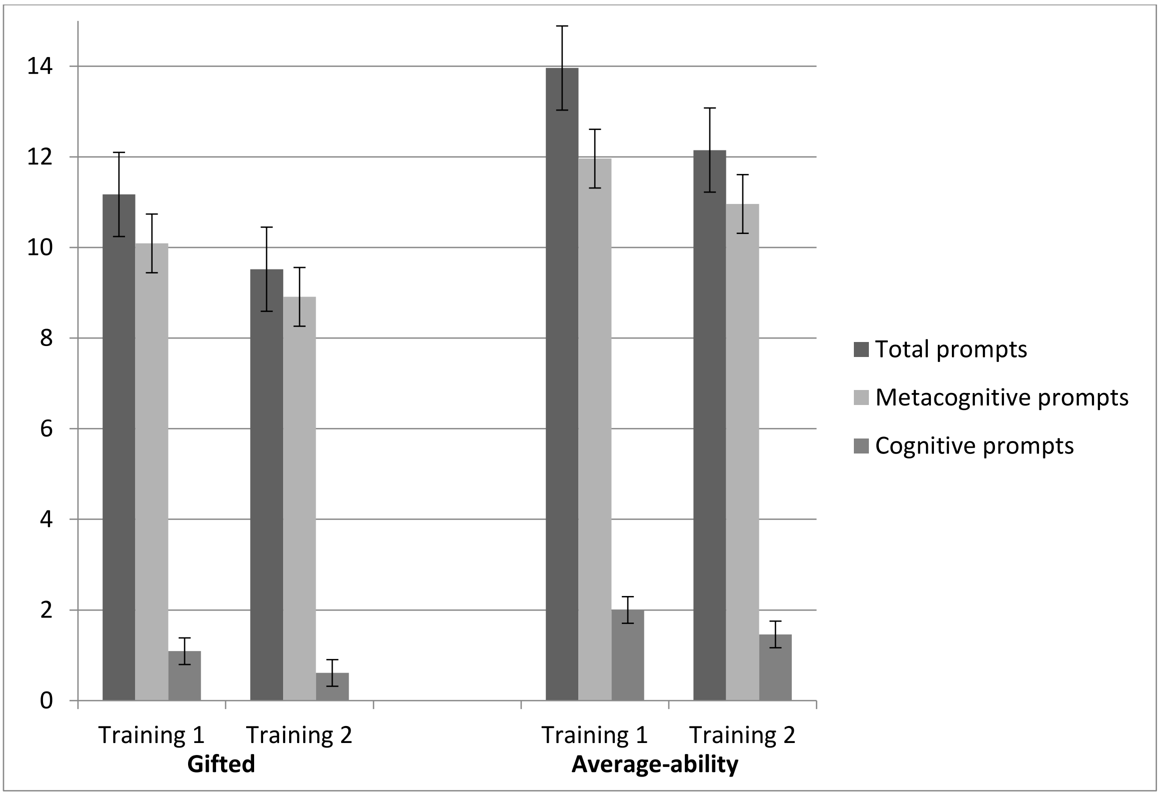

Figure 3. Total Number of Prompts, Metacognitive and Cognitive Prompts per Training Session and Ability Group. Error bars denote standard errors.

3.4. Effect of Planning on Pre-test and Post-test Accuracy and Instructional Needs

3.4.1. Correlation analysis. Prior to conducting hierarchical regression analyses to examine the effect of planning on our static and dynamic measures of analogical reasoning accuracy and instructional needs, Pearson correlations were computed to explore the relationships between these variables. In Table 3, the correlation matrix is shown between planning on the one hand and the accuracy and instructional needs on the other hand, per Ability group and the two conditions.

		Pre-test	Post-test		Total prompts session 1	Total prompts session 2
			Dynamic training	Control		
Planning	Gifted	-.21	-.07	.08	.05	-.07
	Average-ability	-.15	-.19	-.13	.30*	.16

Note. * $p < .05$

Table 3. *Pearson Correlation Matrix between Planning, Pre-test and Post-test accuracy (divided by Condition), and the Total Number of Prompts for both Training Sessions, Divided by Ability Group*

3.4.2. Pre-test. A hierarchical regression model was conducted to examine the potential effects of planning and ability group on pre-test accuracy. Model 1 included the main effect of Planning. This model was significant with 4.6% explained variance (R^2 = .046), $F(1, 146) = 7.09$, $p = .009$. Model 2, in which the main effect of Ability group was entered, was also significant, $F(2, 145) = 30.49$, $p < .001$. A significant improvement in explained variance was found compared with the first model (ΔR^2 = .25, $\Delta F(1, 145) = 41.44$, $p < .001$). Model 3, in which the interaction effect between Ability group and Planning was added, was also significant, $F(3, 144) = 20.26$, $p < .001$, but there was no significant difference in the variance explained as compared with the previous model ($\Delta R^2 = .001$, $\Delta F(1, 144) = .15$, $p = .698$). The parameter estimates of all models are displayed in Table 4. In line with our expectations, planning seemed to predict pre-test accuracy, but only modestly ($b = -.07$, $p = .038$), even when the effect of ability group ($b = 4.42$, $p < .001$) was taken into account. A larger number of steps taken to solve the ToL, corresponding with lower planning abilities, was associated with a lower accuracy score. No moderating effect of Ability group on Planning was found.

	Model 1		Model 2		Model 3	
	B (SE)	Β	*B (SE)*	β	*B (SE)*	β
Constant	7.64 (.34)		7.64 (.29)		7.63 (.30)	
Planning	-.10 (.04)	-.22**	-.07 (.03)	-.15*	-.07 (.03)	-.15*
Ability group			4.45 (.62)	.51***	4.42 (.63)	.50***
Ability group x Planning					.03 (.07)	-.03

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 4. *Hierarchical Multiple Regression Analysis predicting Pre-test Accuracy by Planning and Ability Group*

3.4.3. Post-test. A second hierarchical regression model was computed to examine the potential effects of the graduated prompts training and planning on post-test accuracy, taking into account the moderating effect of ability group. Model 1 included the main effects of Planning and Condition and the interaction effect between Planning and Condition. This model was significant with 8.7% explained variance ($R^2 = .087$), $F(3, 144) = 4.56$, $p = .004$.

Model 2, in which the main effect of Ability group, and the Ability group x Condition interaction effect were entered, was also found to be significant, $F(5, 142) = 10.39$, $p < .001$. A significant improvement in explained variance was found compared with the first model ($\Delta R^2 = .18$, $\Delta F(2, 143) = 17.56$, $p < .001$). Model 3, in which the Ability group x Planning interaction effect was added, was also significant, $F(6, 141) = 8.80$, $p < .001$, but there was no significant difference in the variance explained by the model ($\Delta R^2 = .004$, $\Delta F(1, 141) = .86$, $p = .355$). The parameter estimates of all models are displayed in Table 5. Ability group ($b = 3.62$, $p < .001$), Condition ($b = 2.09$, $p = .001$), and, unexpectedly, Ability group x Condition ($b = -2.78$, $p = .033$), likely the result of a ceiling effect, were significant predictors of post-test accuracy. As expected, and in line with our findings above, being identified as gifted and having been trained seemed to be positively associated with post-test accuracy. However, the predictive value of planning was revealed not to be significant for post-test accuracy when adding the effect of ability group.

	Model 1		Model 2		Model 3	
	B (SE)	β	*B (SE)*	β	*B (SE)*	β
Constant	12.14 (.33)		12.144 (.30)		12.19 (.31)	
Planning	-.07 (.04)	-.16*	-.04 (.03)	-.09	-.04 (.03)	-.09
Condition	2.09 (.67)	.25**	2.11 (.60)	.25**	2.09 (.60)	.25**
Planning x Condition	.022 (.08)	.02	-.02 (.07)	-.02	-.01 (.07)	-.01
Ability group			3.56 (.64)	.40***	3.62 (.65)	.41***
Ability group x Condition			-2.76 (1.29)	-.16*	-2.78 (1.29)	-.16*
Ability group x Planning					.07 (.07)	.07

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 5. *Hierarchical Multiple Regression Analysis predicting Post-test Accuracy by Planning and Ability Group, taking into account the effect of Condition*

3.4.4. Instructional needs. Two final hierarchical regression models were conducted to examine the potential effects of planning and ability group on the total number of prompts children needed during the first and the second training session. The effect of planning and ability group on training session 1 was examined first. Model 1 included the main effect of Planning, and was found to be significant with 6.9% explained variance (R^2 = .069), $F(1, 69)$ = 5.10, p = .027. Model 2, in which the main effect of Ability group was entered, was also found to be significant, $F(2, 68)$ = 4.46, p = .015. However, the improvement in explained variance from Model 1 to Model 2 was bordering on significance only (ΔR^2 = .05, $\Delta F(1, 68)$ = 17.56, p = .061). Model 3, in which the Ability group x Planning interaction effect was added, was also significant, $F(3, 67)$ = 3.35, p = .024, but there was no significant additional improvement in the variance explained by this model (ΔR^2 = .01, ΔF(1, 67) = 1.10, *p* = .298). The parameter estimates of this analysis are displayed in Table 6.

Interestingly, ability group seemed to have the strongest association with the number of prompts children needed at training session 1, but was only marginally significant (*b* = -2.64, *p* = .054), with gifted children needing fewer prompts than their average-ability peers. Planning, moreover, was a significant, but moderate, predictor of the number of prompts, but only in the first (*b* = .15, *p* = .027) and second model (*b* = .14, *p* = .035). In the third model, in which the interaction effect of ability group and planning was added, this predictor was only marginally significant (*b* = .13, *p* = .058). These exploratory results provided a first indication that children who needed more steps solving the ToL items, corresponding with lower planning abilities, needed more prompts.

For training session 2, however, none of the three models were significant in predicting the number of prompts children needed. Model 2, *F*(2, 68) = 3.04, *p* = .054, and 3, *F*(3, 67) = 2.23, *p* = .093, however, were found to be marginally significant. In this analysis, only ability group was a significant predictor of the total number of prompts children needed (*b* = -2.60, *p* = .028).

Table 6.

Hierarchical Multiple Regression Analysis predicting Instructional Needs by Planning and Ability Group

	Training session 1						Training session 2					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
	B (SE)	β	*B (SE)*	β	*B (SE)*	β	*B (SE)*	β	*B (SE)*	β	*B (SE)*	β
Constant	13.03 (.64)		13.02 (.63)		12.96 (.63)		11.29 (.55)		11.27 (.54)		11.24 (.54)	
Planning	.15 (.07)	.26*	.14 (.07)	.25*	.13 (.07)	.22	.06 (.06)	.13	.05 (.06)	.11	.04 (.06)	.09
Ability group			-2.56 (1.34)	-.22	-2.64 (1.35)	-.23*			-2.54 (1.15)	-.26*	-2.60 (1.16)	-.26*
Ability group x Planning					-.16 (.15)	-.21					-.10 (.13)	-.09

Note. * $p < .05$

4. Discussion

The aims of the present study were two-fold. First, we sought to examine the usefulness of a dynamic test of analogical reasoning in obtaining insight into the potential for learning of gifted and average-ability children, focusing both on children's progression after training, and individual instructional needs. Second, we aimed to obtain more insight into the potential unique role that planning plays in unveiling dynamic test outcomes and instructional needs, focusing on potential differences between gifted and average-ability children.

The results of the current study indicated that children who received training between pre-test and post-test progressed more in accurately solved analogies than their peers who were not trained. These results support the notion that testing children dynamically provides more in-depth insight into their potential for learning than testing them statically (Resing & Elliott, 2011; Robinson-Zañartu & Carlson, 2013). It was further found that gifted children, at the group level, achieved higher scores than their average-ability peers. Their progression paths, however, irrespective of whether they were trained, seemed to be parallel to those of their average-ability peers, indicating that these two groups of children benefitted equally from training or practice opportunities.

In relation to children's instructional needs, it was found that children showed a significant decrease in the total number of prompts they needed in training 1 as opposed to training 2. This decrease was also found in relation to the number of metacognitive and cognitive prompts children needed. Although gifted children seemed to need fewer total prompts as well as metacognitive prompts, in accordance with earlier findings (e.g., Kanevsky, 1994), the extent to which their need for prompts decreased from the first to the second training session was equivalent rather than differential, indicating that both groups of children still needed help at the second training session.

With regard to differences between gifted and average-ability children, it must be noted that a ceiling effect may have been encountered by some of the gifted children, who already performed at a high level at the pre-test. If, indeed, there was a ceiling effect, this could account, in part, for the fact that gifted children, as a group, needed fewer prompts. We would then, however, expect these children to show a sharper decrease in their need for prompts than their average-ability peers. This tentative explanation, therefore, requires further research utilizing the test items used in the current study as well as more difficult ones. We cannot rule out that the parallel progression lines were, in part, brought about by regression to the mean. Perhaps in future studies more intensive training, for example by an extra training session, as well as adding a second post-test would lead to more distinct differences between these groups, particularly from the first to the second post-test, if a potential ceiling effect can be controlled for. Moreover, the manner in which children were categorized as gifted or average-ability may have played a role in our research findings. Perhaps if a stricter categorisation system in relation to giftedness identification was used, such as formal giftedness identification, more distinct differences would have been found in the static and dynamic test outcomes of these children.

The second aim of the current study was to obtain more insight into the potential unique association between planning, pre-test, post-test and instructional needs. Our results indicated that planning ability predicted children’s accuracy on the pre-test, but only moderately, with children with weaker planning abilities having lower accuracy scores on the pre-test, which is in line with previous studies (Zook et al., 2004). In addition to planning, as was expected in the light of the findings above, it was found that ability group also significantly predicted pre-test accuracy. Interestingly, however, children’s accuracy scores on the post-test could be predicted only, albeit moderately, by their planning abilities if ability group was not included in the prediction model. It seems that it is, therefore, ability group rather than planning that can predict variability in post-test accuracy, regardless of whether children are trained.

These results seem to suggest that planning only affects children's initial performance in accurately solving analogy items. Apparently, practice and training, making children more expert in a task, help compensate for children's planning skills. Of course, the type of training is of importance when investigating the relationship between planning and dynamic measures of analogical reasoning. In a study by Wiedl et al. (2004), in which adult psychiatric patients were administered a dynamic version of the WCST, for example, planning was associated more strongly with the post-test than with the pre-test. The participants were, however, provided with a short training procedure specifically aimed at improving task orientation, while the procedure utilized in the current study focused on metacognitive and cognitive aspects of the analogical reasoning solving process. As such, different relational patterns would be expected. Future studies could investigate the relationship between dynamic measures of analogical reasoning and planning further by employing a design with two different experimental conditions given a different type of training.

With regard to the relationship between planning and instructional needs, it was found that the instructions children needed in the first training session only could be predicted by their planning abilities, with children with weaker planning skills requiring more prompts, as was also found in earlier research into different executive functions (Stad et al., 2018). These results likely reveal that planning abilities are only initially associated with need for instruction in a training or learning context, and that further learning experiences counter the influence of planning.

No differences were found between gifted and average-ability children in relation to the extent to which planning influenced test scores and their need for instructions. These findings suggest that executive functioning seems to play a similar role in the process of solving analogy items of the two groups of children, despite the idea that gifted children excel in executive functions (e.g., Arffa, 2007). More importantly in this respect, however, individual differences were found in pre-test and post-test accuracy, instructional needs, as well as in performance on the ToL within the two ability groups. It seems worthwhile to focus on these individual differences in future studies, aiming to further uncover the relationship between executive functions and testing outcomes.

In addition to the potential ceiling effect and the giftedness identification procedure used in the current study, some additional limitations were encountered. All in all, the study included a relatively low number of participants, with the four subgroups being relatively small, which of course has an effect on the generalisability of the study. Moreover, the current study analysed one measure of planning, providing insight into one aspect of planning ability. It can, furthermore, not be discounted that the ToL might not only measure planning, but also, more generally, visual-spatial working memory (Gilhooly, Wynn, Phillips, Logie, & Sala, 2002). In future studies, therefore, it is advisable to examine different aspects of children's planning ability utilizing several different tasks as well as different measures, such as accuracy (Yang, Chen, Wang, & Zhu, 2017), completion time or the time it takes children to perform the first move (Vakil & Heled, 2016). Investigating both proficiency and efficiency in planning measures will potentially provide a more in-depth insight into the role of children's planning abilities in static and dynamic test outcomes.

In conclusion, the current study showed that dynamic testing is a useful instrument for assessing the cognitive potential of children of different ability levels. Although in practice, it is sometimes assumed that gifted children can manage their learning on their own (e.g., De Boer, Minnaert & Kamphof, 2013), the results of the current study clearly show that these children also profit from additional instruction and help in unveiling their cognitive potential. Analysis of these children's instructional needs further indicates that gifted children show large individual differences in their need for instructions, which is relevant in relation to tailoring education to individual needs. Educational psychologists could analyse the outcomes of a dynamic test, such as children's progression in learning and needs for instruction, and translate these to hands-on information relevant for teachers in drawing up individual educational plans for children. Planning, moreover, seems to be associated with initial inductive reasoning performance, as well as with the need for instruction, particularly in the initial stages of giving instructions and help for new tasks, suggesting that differentiation in the number and type of instructions given to children with different planning abilities is in particular important in the first stages of giving instructions for solving new tasks in the classroom.

The fact that both the gifted and average-ability children in our sample of participants portrayed individual differences in both the extent to which they improved in analogical reasoning accuracy, as well as the instructions they needed to demonstrate learning are indications that usage of dynamic tests is useful for children of different ability levels. An important implication of our findings is that traditional, static tests do not always provide a good indication of the level of children's potential for learning, as witnessed by the differences in the extent to which children showed progression of learning after training (e.g. Elliott et al., 2018). More importantly, these results imply that abilities are developmental in nature, which cannot be fully captured by a single-occasion administered test, such as traditional intelligence and school aptitude tests. Information arising from such test outcomes should therefore be interpreted with caution, by an educational psychologist, before it is shared with educational staff, parents, and children. If the cognitive abilities of children are questioned, it is advised to use a dynamic in addition to or instead of a static test. Not only will these tests most likely provide a more complete picture of children's cognitive potential as well as an overview of their instructional needs, they also seem to compensate for the effect of executive functions on performance, thereby providing a less biased insight into what children are really capable of.

5. References

Arffa, S. (2007). The relationship of intelligence to executive function and non-executive function measures in a sample of average, above average, and gifted youth. *Archives of Clinical Neuropsychology*, *22*, 969-978. doi: 10.1016/j.acn.2007.08.001

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*, 205–228. doi:10.1080/87565640801982312

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. doi:10.1016/j.lindif.2010.11.025

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82-109). New York, NY: Guilford Press.

Cormier, P., Carlson, J. S., & Das, J. P. (1990). Planning ability and cognitive performance: The compensatory effects of a dynamic assessment approach. *Learning and Individual Differences, 2*, 437–449. doi:10.1016/1041-6080(90)90004-z

D'Antuono, G., La Torre, F. R., Marin, D., Antonucci, G., Piccardi, L., & Guariglia, C. (2017). Role of working memory, inhibition, and fluid intelligence in the performance of the Tower of London task. *Applied Neuropsychology: Adult*, *24*, 548-558. doi:10.1080/23279095.2016.1225071

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:10.1177/0162353212471622

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, *33*, 261–268. doi:10.1016/0028-3932(94)00124-8

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

Gilhooly, K. J., Wynn, V., Phillips, L. H., Logie, R. H., & Sala, S. D. (2002). Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual differences approach. *Thinking & Reasoning*, *8*, 165–178. doi:10.1080/13546780244000006

Gligorovic, M., & Buha, N. (2016). Influence of inhibitory control on planning abilities in children with mild intellectual disability. *Specijalna Edukacija i Rehabilitacija*, *15*, 287–304. doi:10.5937/specedreh15-12259

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

Kanevsky, L. S. (1994). Exploring the implications of dynamic and static assessments for gifted education. *Exceptionality Education Canada*, *4*, 77–98.

Kar, B. C., Dash, U. N., Das, J. P., & Carlson, J. (1993). Two experiments on the dynamic assessment of planning. *Learning and Individual Differences*, *5*, 13–29. doi:10.1016/1041-6080(93)90023-l

McCormack, T., & Atance, C. M. (2011). Planning in young children: A review and synthesis. *Developmental Review*, *31*, 1–31. doi:10.1016/j.dr.2011.02.002

Monette, S., Bigras, M., & Guay, M.-C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *Journal of Experimental Child Psychology*, *109*, 158–173. doi:10.1016/j.jecp.2011.01.008

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, *222*, 250-259. doi:10.1016/j.jneumeth.2013.10.024

Passig, D., Tzuriel, D., & Eshel-Kedmi, G. (2016). Improving children's cognitive modifiability by dynamic assessment in 3D Immersive Virtual Reality environments. *Computers & Education*, *95*, 296-308. doi: 10.1016/j.compedu.2016.01.009

Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory*, *7*, 209–231. doi:10.1080/741944066

Piper, B. J., Mueller, S. T., Geerken, A. R., Dixon, K. L., Kroliczak, G., Olsen, R. H. J., & Miller, J. K. (2015). Reliability and validity of neurobehavioral function on the Psychology Experimental Building Language test battery in young adults. *PeerJ*, *3*, e1460. doi:10.7717/peerj.1460

Raven, J., Raven, J. C., & Court, J. H. (2000). *Standard progressive matrices*. Oxford, England: Psychology Press.

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. doi:10.1348/2044-8279.002006

Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science*, *24*, 87-92. doi: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. doi: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, Vol. 3.* (pp. 149-167). Washington, DC: American Psychological Association. doi:10.1037/14049-007

Ropovik, I. (2014). Do executive functions predict the ability to learn problem-solving principles? *Intelligence, 44*, 64–74. doi:10.1016/j.intell.2014.03.002

Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H., & Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology*, *15*, 232–246. doi:10.1080/09297040802220029

Stad, F. E., Wiedl, K. H., Vogelaar, B., Bakker, M., & Resing, W. C. M. (2018). The role of cognitive flexibility in young children's potential for learning under dynamic testing conditions. *European Journal of Psychology of Education*. doi:10.1007/s10212-018-0379-8

Stevenson, C. E., Heiser, W. J., & Resing, W. C. M. (2013). Working memory as a moderator of training and transfer of analogical reasoning in children. *Contemporary Educational Psychology*, *38*, 159–169. doi:10.1016/j.cedpsych.2013.02.001

Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology*, *98*, 265–281. doi:10.1037/0022-0663.98.2.265

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. doi:10.1016/j.lindif.2007.07.007

Unterrainer, J. M., Rahm, B., Kaller, C. P., Leonhart, R., Quiske, K., Hoppe-Seyler, K., ... & Halsband, U. (2004). Planning abilities and the Tower of London: Is this task measuring a discrete cognitive function?. *Journal of Clinical and Axperimental Neuropsychology*, 26, 846-856. doi: 10.1080/13803390490509574

Vakil, E., & Heled, E. (2016). The effect of constant versus varied training on transfer in a cognitive skill learning task: The case of the Tower of Hanoi Puzzle. *Learning and Individual Differences*, *47*, 207-214. doi:10.1016/j.lindif.2016.02.009

Vogelaar, B., Bakker, M., Hoogeveen, L., & Resing, W. C. M. (2017). Dynamic testing of gifted and average-ability children's analogy problem solving: Does executive functioning play a role? *Psychology in the Schools*, *54*, 837-851. doi: 10.1002/pits.22032

Wiedl, K. H., Schöttke, H., Green, M. F., & Nuechterlein, K. H. (2004). Dynamic testing in schizophrenia: Does training change the construct validity of a Test? *Schizophrenia Bulletin*, *30*, 703–710. doi:10.1093/oxfordjournals.schbul.a007124

Yang, X.H., Chen, Z., Wang, Z.H., & Zhu, L.Q. (2017). The relations between television exposure and executive function in Chinese preschoolers: The moderated role of parental mediation behaviors. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.01833

Zook, N. A., Davalos, D. B., DeLosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and Cognition*, *56*, 286–292. doi:10.1016/j.bandc.2004.07.003