



Universiteit
Leiden
The Netherlands

Dynamic testing and excellence: unfolding potential

Vogelaar, B.

Citation

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

Version: Not Applicable (or Unknown)

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

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

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

Cover Page



Universiteit Leiden



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

Author: Vogelaar, B.

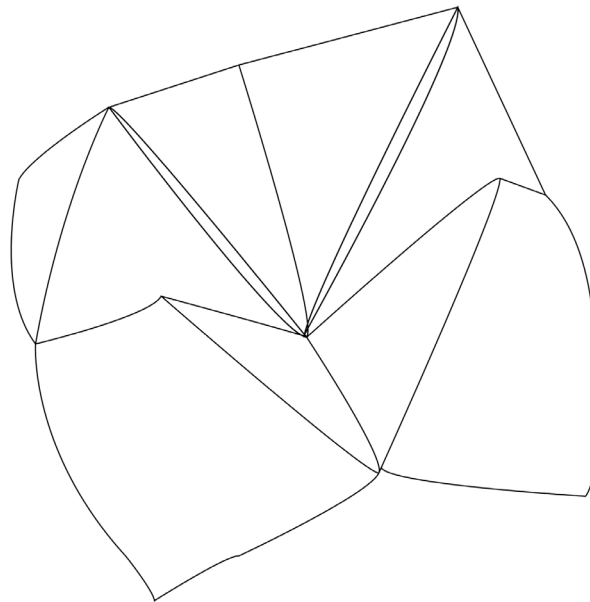
Title: Dynamic testing and excellence: unfolding potential

Issue Date: 2017-01-18

CHAPTER 3

Dynamic testing and test anxiety amongst gifted and average-ability children

3



Bart Vogelaar
Merel Bakker
Julian G. Elliott
Wilma C. M. Resing

Vogelaar, B., Bakker, M., Elliott, J. G., & Resing, W. C. M. (under revision).
Dynamic testing and test anxiety amongst gifted and average-ability children.
[British Journal of Educational Psychology]

Abstract

Background. Dynamic testing has been proposed as a testing approach that is less disadvantageous for children who may be potentially subject to bias when undertaking conventional assessments. For example, those who encounter high levels of test anxiety, or who are unfamiliar with standardised test procedures, may fail to demonstrate their true potential or capabilities. While dynamic testing has proven particularly useful for special groups of children, it has rarely been used with gifted children.

Aim. We investigated whether it would be useful to conduct a dynamic test to measure the cognitive abilities of intellectually gifted children. We also investigated whether test anxiety scores would be related to a progression in the children's test scores after dynamic training. **Sample.** Participants were 113 children aged between 7 and 8 years from several schools in the western part of the Netherlands. The children were categorised as either gifted or average-ability, and split into an unguided practice or a dynamic testing condition.

Methods. The study employed a pre-test-training-posttest design. Using Linear Mixed Modeling analysis with a multilevel approach we inspected the growth trajectories of children in the various conditions, and examined the impact of ability and test anxiety on progression and training benefits.

Results and Conclusions. Dynamic testing proved to be successful in improving the scores of the children, although no differences in training benefits were found between gifted and average-ability children. Test anxiety was shown to influence the children's rate of change across all test sessions, and their improvement in performance accuracy after dynamic training.

3.1. Introduction

Over the last few decades, the notion that gifted and talented children might need special assistance in their learning has become increasingly acknowledged. For a long time, it has been a commonly held belief that this group of children could manage classroom learning on their own. Fortunately, with greater recognition that the notion of inclusive education should apply to all children, increasing attention is being paid to the educational needs of gifted and talented children (De Boer, Minnaert, & Kamphof, 2013).

Formal assessment of intellectual giftedness typically involves the use of conventional, static assessments of intelligence or school achievement (Kline, 2001). These tests, however, have been shown to be disadvantageous for certain groups of children (Haywood & Lidz, 2007), such as those who experience test anxiety (Meijer, 1996, 2001). In contrast to static, conventional tests, dynamic tests incorporate feedback and instruction into the testing procedure (Elliott, Grigorenko, & Resing, 2010), and are considered to tap into individual children's potential for learning (Sternberg & Grigorenko, 2009). In addition, the literature on dynamic testing has indicated that static tests may underestimate the cognitive potential of socially or educationally disadvantaged children. Examples include ethnic minority, learning disabled, or those who have not had access to educationally stimulating environments (Grigorenko & Sternberg, 1998; Haywood & Lidz, 2007; Robinson-Zañartu & Carlson, 2013). In contrast, dynamic tests are considered to have less test bias towards such children (Elliott, 2003).

The focus of our current study was two-fold. We investigated whether it would be useful to conduct a dynamic test in order to measure the cognitive abilities of intellectually gifted children. In addition, we investigated whether test anxiety scores would be related to progression in test scores after dynamic training.

Dynamic testing

Dynamic testing has been described as an umbrella concept used to denote a form of testing that is focused on a child's potential for learning, rather than as a measure of their previous learning (Sternberg & Grigorenko, 2002). The most frequently used application of dynamic testing is the pre-test-training-post-test design, which enables structured measurement of the learning progression of an individual child (Sternberg & Grigorenko, 2009). In such a design, different intervention, or training, approaches can be implemented, an example of which is the graduated prompts technique (Campione & Brown, 1987). This technique involves a hierarchically structured approach in which children receive a graduated series of prompts that become more specific in relation to the solution

of the task with each new prompt. In the current study, we used a dynamic approach (Resing, 2000) to examine progression in analogical problem-solving. Our participant sample consisted of seven and eight year old children who were split into gifted and average-ability groups. Analogical reasoning, a subtype of inductive reasoning, is considered to play a central role in cognitive development (Klauer & Phye, 2008; Pellegrino & Glaser, 1982). Empirical studies have shown that this ability develops significantly in young primary school children (e.g., Tunteler & Resing, 2007).

The large majority of studies into dynamic testing have focused on the special populations mentioned above. Far more scarce are studies applying dynamic testing to children who have the potential to excel (although, see Lidz & Elliott, 2006). Most dynamic testing studies involving talented or gifted children have focused upon children who are considered to suffer bias in conventional test settings, such as those with a low SES (e.g., Frasier & Passow, 1994), or ethnic minorities (e.g., Lidz & Macrine, 2001). Empirical studies indicate that the cognitive advantage of gifted and talented children is expressed by a more extensive zone of proximal development (e.g., Calero, García-Martín, & Robles, 2011). Such studies show they learn new skills faster, and have an advantage in generalising knowledge (e.g., Kanevsky, 2000). The role that test anxiety potentially plays amongst this group of learners when they are dynamically tested rather than in a conventional static fashion has not been studied before, and this was a key aim of the current study.

Test anxiety

Test anxiety has been described as a negative emotional or cognitive response to situations in which performance is being measured or assessed (Cassady & Johnson, 2002). It is comprised of two dimensions: a cognitive and an emotional component (McDonald, 2001). The cognitive component of test anxiety has been described as consisting of worrying and negative thoughts that are unwanted, uncontrollable and aversive, and which lead to emotional discomfort (Davey, 1994). This component can often occur before, during and after an evaluation or an assessment (Cassady & Johnson, 2002). Some empirical studies have suggested that the prevalence of test anxiety may be lower amongst children with the potential to excel than amongst children with average-ability (Davis & Connell, 1985; Wooding & Bingham, 1988; Zeidner & Schleyer, 1999). It has been hypothesised that this may be due to these children having higher intellectual coping resources that lead them to cope better in stressful academic situations (Zeidner & Shani-Zinovich, 2011).

The consequences of high levels of test anxiety are well-known, ranging from

underperformance on standardised tests, allocation to lower performing groups in school to dropping out of school altogether (Everson, Millsap, & Rodriguez, 1991; Hancock, 2001; Sub & Prabha, 2003). A variety of research has shown that students who experience high levels of test anxiety perform significantly lower on school tests, and are found to have a lower grade point average (e.g., Segool, Carlson, Goforth, Von der Embse, & Barterian, 2013). In addition, some studies have found that test anxiety may have a negative impact on intelligence test performance (e.g., Meijer, 2001; Morris & Liebert, 1969) with some authors finding a moderate negative correlation of -0.2 between test anxiety and static measures of intelligence (Zeidner, 1998).

Whereas the relationship between test anxiety and static intelligence and educational tests has been heavily researched, there are only few studies investigating the association between test anxiety and performance on dynamic tests. These studies do, nevertheless, support the expectation that testing dynamically rather than statically is advantageous for children who experience test anxiety. Meijer (1996, 2001), for example, found that amongst adolescent learners, dynamic mathematics tests showed less bias towards children experiencing test anxiety than conventional, static mathematics tests. A study by Bethge, Carlson, and Wiedl (1982) revealed that amongst third grade children, test anxiety seems to be diminished when children's analogical reasoning ability was assessed dynamically. No study, however, has investigated the relationship between test anxiety and test performance in a dynamic test context, on the one hand, and potential differences between gifted and average-ability children, on the other.

The current study

Our first task was to investigate the potential effects of dynamic testing for gifted and average-ability children. We compared their progression paths from pre-test to post-test in both a dynamic training and an unguided practice group. We (1) expected a main effect of condition, and hypothesised that children who received dynamic testing (which incorporated a short training session) would show more progression in analogical reasoning than children who received unguided practice only (Resing, 2000; Stevenson, Hickendorff, Resing, Heiser, & de Boeck, 2013). In addition, we focused on any potential differences between gifted and average-ability children. We expected an interaction between condition and ability category, and hypothesised (1a) that the dynamically trained gifted children would show more advanced progression paths in analogical reasoning than their dynamically trained average-ability peers (Calero et al., 2011; Kanevsky, 2000), and (1b) that the gifted children in the unguided practice

condition would also show more progression than their average-ability peers in the unguided practice condition (Calero et al., 2011).

Our second aim was to provide insight into the association between test anxiety and progression in test performance after dynamic testing. First of all, we expected that test anxiety would influence the level of accuracy scores of analogical reasoning. Given that in prior research with adolescent learners, dynamic testing has indicated lower test anxiety bias than static testing (Meijer, 1996, 2001), we expected a significant interaction between test anxiety and condition. In relation to the effect of training, we expected to find a differential effect of dynamic training on children with different levels of test anxiety. More specifically, we hypothesised (2a) that children with higher test anxiety scores would benefit more from training than children with lower test anxiety scores. Focusing on differences between the gifted and average-ability children, we also expected a significant interaction between condition, test anxiety and ability category. We further hypothesised (2b) that the progression paths of average-ability children with higher levels of test anxiety would be steeper than their gifted peers with higher levels of test anxiety (Zeidner & Shani-Zinovich, 2011).

3.2. Method

Participants

Study participants were 113 children, 54 boys and 59 girls, ranging in age from 7 years and 1 month to 8 years and 9 months ($M=7.91$ in years, $SD=6.40$ in months). All the children were born in the Netherlands, and attended mainstream primary schools or were enrolled in special settings for gifted and talented children in the western part of the Netherlands. In this country, intelligence testing is not standard practice in primary schools and placement into gifted or talented programmes is often based on the qualitative judgements of parents and teachers. Schools participated on a voluntary basis. Gifted children were over-sampled and identified on the basis of a qualitative judgment of parents and teachers regarding their giftedness. Additionally, all of the children in our gifted sample each scored at, or above the 90th percentile on the Raven's Progressive Matrices Test (Raven, 1981). Written permission from parents and schools to participate in the study was obtained for each child. Six children dropped out in the course of the study, as they did not participate in each test session. Their data were not included in the analyses.

Design

The study used a three-session (pre-test 1, pre-test 2, post-test) repeated measures randomised blocking design with two treatment conditions: dynamic training versus unguided practice (see Table 1). Half of the children received a dynamic training session between pre-test 2 and post-test, whereas the other half

of the children, allocated to the unguided practice condition, received a dot-to-dot control task. Before the actual study commenced, prior to pre-test 1, the Raven Progressive Matrices Test (Raven, 1981) was administered to allocate the children to the various conditions. Children with Raven percentile scores of at least the 90th percentile were allocated to the "gifted" condition; the other children to the average-ability condition. Further, Raven scores were used to ensure that any differences in initial reasoning ability were as small as possible across the children in the dynamic training and unguided practice conditions. Pairs of children with equal scores (blocking) were randomly assigned to the dynamic testing or unguided practice condition, resulting in four subgroups: gifted dynamic training (N=22), gifted unguided practice (N=23), average-ability dynamic training (N=31) and average-ability unguided practice (N=37).

Our design included pre-test sessions 1 and 2 in order to enable comparison between static and dynamic progression. During the pre-test sessions and the post-test, the children were provided with only short, general instructions and were not given any feedback. After the post-test, all children were asked to complete the Children's Test Anxiety Scale (CTAS), a domain-general self-report questionnaire measuring test anxiety amongst children in grades 3-6 of elementary school. Administration of the instruments in the three sessions and the dynamic training each took approximately 20-30 minutes.

Table 1. Overview of the design

Groups		Pre dynamic testing	Dynamic/Static test				Post dynamic testing
Condition		Raven	Pre-test 1	Pre-test 2	Dynamic training	Post-test	CTAS
Dynamic training (N=53)	Gifted (22)	X	X	X	Dynamic training	X	X
	Average-ability (31)						
Unguided practice (N=60)	Gifted (23)	X	X	X	Dot-to-dot control task	X	X
	Average-ability (27)						

Materials

Raven. The Raven Progressive Matrices Test (Raven, 1981) was administered to all children as a blocking instrument. The Raven is a non-verbal intelligence test measuring fluid intelligence by means of multiple choice figural analogies. In our sample of participants, the internal consistency of the Raven accuracy scores was found to be high, as measured by Cronbach's α of .94.

Children's Test Anxiety Scale (CTAS). To measure test anxiety in children, a Dutch translation of the Children's Test Anxiety Scale (CTAS) was used (Wren & Benson, 2004). The CTAS is a 30 item self-report questionnaire for school children in grades 3 through 6 that utilises a 5-point Likert scale. Here, children were asked to answer statements on three dimensions (their thoughts, autonomic reactions, and behaviour) measured by the questionnaire, when taking tests. The internal consistency of the CTAS was found to be high in our sample of participants (Cronbach's $\alpha = .92$).

Dynamic test of analogical reasoning. The dynamic test used in the present study consisted of open-ended series of geometric analogies, of varying difficulty, of the type $A:B::C:D$, assumed to measure inductive reasoning (Barnett & Ceci, 2002). The pre-tests and the post-test, parallel sessions, included 20 analogy items of various difficulty, originally created by Hosenfeld, Van den Boom, and Resing (1997), and adapted by Tunteler, Pronk, and Resing (2008). Six basic geometrical shapes were used in each analogy item: squares, triangles, hexagons, pentagons, circles, and ovals. Each analogy item contained five possible transformations: changing position, adding or subtracting an element, changing size, halving, and doubling (Hosenfeld et al., 1997). The test was administered as an open-ended paper-and-pencil test and the children had to draw their own answers. Figure 1 shows an example of a difficult item.

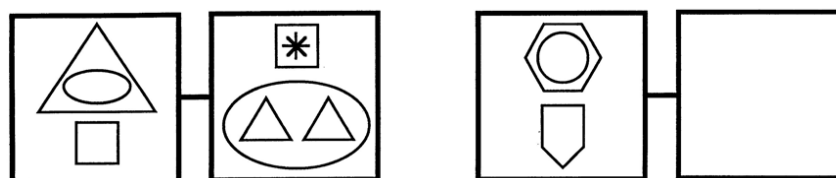


Figure 1. Example of a difficult analogy item.

Pre-tests and post-test. The two pre-tests and the post-test each contained 20 items of varying difficulty. Participants received minimal instructions only; they were instructed to solve puzzles with different shapes. Each puzzle had three boxes that were filled, and an empty one. The tester then asked the child which

shapes had to be drawn in the fourth box in order to solve the puzzle. Pre-test 1 was found to have high internal consistency (Cronbach's $\alpha = .94$).

Dynamic training. The dynamic training session consisted of 10 new geometric analogy problems. The training session employed graduated prompts techniques that have been employed in earlier studies (e.g. Resing & Elliott, 2011). These involve the provision of a number of prompts when the child makes an error. All prompts were administered hierarchically: starting with two very general metacognitive prompts followed by two concrete cognitive prompts tailor-made for each item. As each new prompt progressively became more specific, this procedure enabled the measurement of the child's use of differing degrees of help. The training session consisted of five steps in total. Prompts were only administered after indication that a child could not solve the analogy independently. At each step, children were asked to draw the solution of the analogy, and check whether their solution was correct. If a child had not solved the analogy after the fourth prompt had been administered, the tester modelled the correct answer. After responding, participants were asked to explain why they thought their answer was correct. Finally, the tester provided a correct self-explanation. Figure 2 consists of a flowchart of the training procedure.

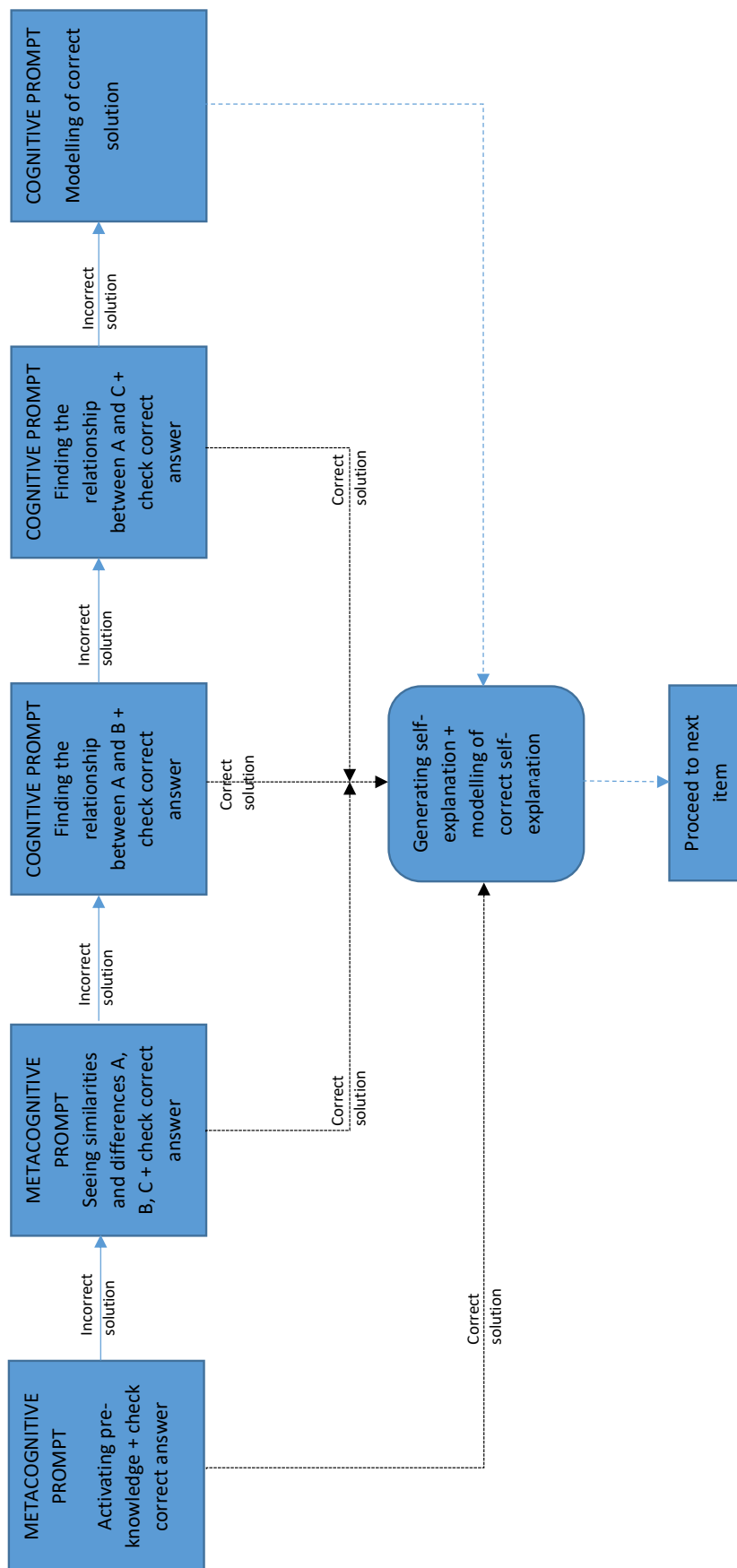


Figure 2. Flowchart of the graduated prompts training protocol.

General procedure

Children in the current study were tested once a week over a period of five consecutive weeks. All tests and questionnaires were administered following standard, protocolled instruction. At the beginning of the pre-tests, the training sessions and the post-tests, children were given a sheet containing the six geometrical shapes used in the analogies, and were asked to name each shape. Then, the tester asked the child to draw the shapes below the printed models, staying as close to the original as possible (Tunteler et al., 2008). This procedure was supposed to help activate the children's prior knowledge, ensured that the tester and child used the same terms for the geometric shapes used in the analogy, and facilitated the scoring procedure.

Analysis

We considered the current study to be comprised of multilevel data, where the repeated measurements were nested within children (Hox, 2002, 2010; Kreft & De Leeuw, 2007; Snijders & Bosker, 1999; Van der Leeden, 1998). Multilevel analysis allowed us to model the training effect and the effects of repeated practice separately, and across sessions. This enabled us to investigate the systematic variation between these trajectories as a function of our experimental treatment and predictor variables (Van der Leeden, 1998).

Linear Mixed Modeling analysis, with a multilevel approach (with the lme4 package; Bates, Mächler, Bolker, & Walker, 2015), was used to inspect the growth trajectories of children in the various conditions. Level 1 represented the repeated measurements of the number of correct items within children, and level 2 represented the variability between children. We could therefore model the average growth trajectories of various groups of children (Hox, 2002, 2010).

The models were fitted in R (R Development Core Team, 2014), and the parameters of the models were estimated with full maximum likelihood. We included the predictor variables (time-constant and time-varying variables) in the model in the order of our hypotheses. First, an unconditional means model was carried out that included a random intercept. Next, we included the linear effect of time in the unconditional growth model. These models were carried out to analyse the variance in the number of correct analogies between children and over time within children. The subsequent, conditional models included the following predictors: condition, ability category, and test anxiety. We centred the time-invariant predictor Test anxiety by subtracting the sample mean from each observed value. Recentring was applied in order to improve interpretation (Singer & Willett, 2003). Likelihood ratio (LR) tests (Chi-square distributed) and model-fit indices (the Akaike Information Criterion (AIC), and the Bayesian

Information Criterion (BIC)) were examined to assess the difference in model fit of the successive models. The AIC and BIC are two ad hoc criteria that are based on the log likelihood statistic. Both indices were used for model selection by comparing the relative goodness-of-fit of models (Singer & Willett, 2003).

3.3. Results

Before using the multilevel models to examine our research questions, one-way analyses of variance were conducted to evaluate possible differences between the two experimental conditions and ability subgroups, respectively, in relation to children's level of inductive reasoning prior to the experiment, age, pre-test 1 accuracy and test anxiety scores. The total Raven scores, as a measure of children's initial level of inductive reasoning, pre-test 1 accuracy scores, test anxiety, and age in months were used as dependent variables and Condition with two levels (dynamic training versus unguided practice) as the independent variable. No significant differences were found in Raven scores ($p=.73$), pre-test 1 accuracy scores, ($p=.31$), test anxiety ($p=.32$) nor in age ($p=.39$) between the dynamic training and unguided practice groups. For the gifted and average-ability children, no differences were found concerning test anxiety ($p=.45$), and age ($p=.31$). As expected, the gifted children outperformed their peers on both the Raven scores ($M=44.20$, $SD=3.97$), and the pre-test 1 accuracy scores ($M=12.69$, $SD=4.42$) (the difference is statistically significant for both measures, $p<.001$). Descriptive statistics are provided in Table 2.

In addition, as part of our preliminary analysis, separate Pearson's product-moment correlations were calculated for each subgroup to investigate potential differences in the relationship between pre-test 1 and post-test accuracy scores in the two conditions. The correlations showed that the association between the pre-test 1 and post-test accuracy was stronger for the children in the unguided practice condition ($r=.83$, $p<.001$) than the children who were dynamically trained ($r=.61$, $p<.001$). This provided a preliminary indication of the validity of the dynamic test.

Table 2. Mean scores and standard deviations of Raven scores, pre-test 1, pre-test 2 and post-test accuracy scores divided by ability category and condition

		Gifted		Average-ability	
		Dynamic training	Unguided practice	Dynamic training	Unguided practice
N		22	23	31	37
Raven	M	43.82	44.57	34.55	33.78
	SD	4.22	3.78	5.53	6.47
Pre-test 1	M	12.00	13.35	9.65	9.22
	SD	5.26	3.41	4.44	4.82
Pre-test 2	M	15.50	17.09	13.84	13.11
	SD	5.63	2.80	4.77	5.95
Post-test	M	17.91	17.04	16.61	12.62
	SD	3.22	2.50	2.86	6.05
CTAS	M	49.82	54.52	53.58	55.43
	SD	12.90	17.44	14.55	18.79

Growth curve analyses (MLA) were used to model growth for the outcome variable, the number of correct analogies. The obtained estimates and fit indices of the models are provided in Table 3. The unconditional means model (Model 1) showed a significant fixed effect of the intercept ($p < .001$). The intra-class correlation coefficient (ICC) indicated that 55.23% of the total variation in the analogy scores was attributable to differences between children. We included our time predictor into the level-1 sub-model in order to explain the remaining within-child variance (12.57).

The effect of Time was included in Model 2 (the unconditional growth model). The children, on average, increased their reasoning accuracy across sessions, as indicated by a significant fixed effect of time (2.47, $p < .001$). We found a negative covariance (-0.40) between the slope and intercept, which revealed that children with lower initial analogy scores generally showed higher rates of progression across test sessions than children with higher initial scores. Inspection of the variance components revealed large remaining variance in the number of correct analogies both between, and within, children. The R^2 value of 0.53 indicated that 53.3% of the within-person variation in reasoning accuracy was accounted for by the linear effect of time. In Model 3 we included the main effect of Condition. We used a likelihood ratio test (LRT) to assess whether model fit improved. The inclusion of Condition led, as expected, to a significant

improvement in model fit ($X^2(1)=7.00$, $p<.001$). The estimated rate of change for an average participant of the repeated practice group was 2.12, indicating that the children generally increased their number of correct analogies across sessions. The positive fixed effect (1.46) for condition (training versus unguided practice) revealed that there was an effect of the dynamic training session on children's progression in the number of correct analogies. As shown in Table 2, and in accordance with our expectation, the children who received a dynamic training showed greater improvement in accuracy scores from pre-test 2 to post-test than the children in the unguided practice condition.

The inclusion of the main effect of Ability category in Model 4 led to an improvement in model fit ($X^2(1)=13.25$, $p<.001$). The significant main effect revealed that children's Ability, gifted versus average-ability, influenced their analogical performance at the first test session. The positive fixed effect of Ability (3.00) showed that children obtained, on average, higher pre-test 1 scores than their average-ability age-mates. However, the non-significant interaction of Ability and Time in Model 5 revealed that Ability did not influence the rate of change in children's reasoning performance ($X^2(1)=0.19$, $p=0.66$). We can conclude that the gifted children who repeatedly practised solving the analogies showed no more progression in accuracy than average-ability peers who also repeatedly practiced.

In Model 6 we included the interaction effect of Ability and Condition to examine whether the dynamic graduated prompts training intervention had a differential effect on the performance of gifted and average-ability children. Model fit did not improve ($X^2(1)=1.49$, $p=.22$). The non-significant interaction effect of Ability and Condition showed, contrary to our expectations, that no significant differences existed in the benefits of dynamic training for the two ability categories.

Model 7 included the main effect of Test anxiety. We found a non-significant improvement in model fit ($X^2(1)=2.26$, $p=.13$). Model 8 however included the interaction effect of Test anxiety and Time. The inclusion of this interaction term led to an improved model ($X^2(1)=10.80$, $p<.005$), indicating that test anxiety influenced the children's rate of improvement in the number of correct analogies. Children with higher test anxiety improved more across test sessions than those experiencing lower levels of test anxiety. The significant interaction effect of Test anxiety x Condition in Model 9 indicated that, as expected, Test anxiety impacted upon the dynamic training benefits of children in the training condition ($X^2(1)=6.49$, $p=.011$). More specifically, children who scored higher on test anxiety improved more from pre-test 2 to post-test. The three-way interaction of Ability category x Condition x Test anxiety in Model 10, however, did not improve model

fit ($X^2(1)=0.97, =0.33$). The progression paths in accuracy scores of gifted children and average-ability peers were, contrary to our expectations, influenced similarly by test anxiety.

Table 3. Results of the fitted multilevel models for the number of correct analogies

Model	Estimate (SE)	Deviance	AIC	BIC
1. Intercept only	13.65(0.42)**	1996.4	2002.4	2013.9
2. Time	2.47(0.18)**	1844.7	1856.7	1879.7
3. Condition	1.46(0.50)*	1837.7	1851.7	1878.5
4. Ability category	3.00(0.80)**	1824.5	1840.5	1871.1
5. Ability category x Time	-0.15(0.34)	1824.3	1842.3	1876.7
6. Ability category x condition	-1.02(0.83)	1823.0	1841.0	1875.4
7. Test anxiety	-0.04(0.02)	1822.2	1840.2	1874.7
8. Test anxiety x Time	0.03(0.01)*	1813.7	1833.7	1872.0
9. Test anxiety x Condition	0.09(0.03)*	1807.2	1829.2	1871.3
10. Ability category x condition x Test anxiety	-0.06(0.06)	1806.2	1830.2	1876.1

Note. Significance: ** $p < .001$, * $p < .05$. The deviance, AIC, and BIC statistics were used to compare the relative goodness-of-fit of the successive models.

After running the multilevel analysis, Model 9 proved to be the best fitting model based on the LRT, AIC, and BIC values. We can conclude that the dynamic sessions were, as expected, successful in improving the scores of the children. In contrast to what we hypothesised, we found no difference in dynamic training benefits between gifted and average-ability children. There was also no effect of Ability category on the accuracy progression of gifted and average-ability children in the unguided practice condition. In line with our hypotheses, test anxiety was shown to influence the children's rate of change across all test sessions and their improvement in accuracy after dynamic training. Lastly, and counter to our expectations, test anxiety did not have less influence on the progression paths of gifted children in comparison with average-ability children.

3.4. Discussion

The current study sought to investigate the potentially different influence of dynamic testing on the performance of average-ability, and gifted learners. In accordance with our expectations, the pre-test-post-test correlations of the children in the two experimental conditions differed. In addition, the results revealed that children who were trained dynamically showed more advanced

progression paths from pre-test to post-test in analogical reasoning than the children who had unguided practice experiences only. This finding lends support to the claims of many researchers that dynamic testing can offer a more complete picture of children's cognitive capacities than conventional static approaches (e.g., Elliott, 2003; Elliott et al., 2010; Sternberg & Grigorenko, 2002). By focusing on what children can learn within a short time-frame, rather than on what children have already learned, dynamic testing appears to unveil children's potential for learning (Robinson-Zañartu & Carlson, 2013), which, as shown in the current investigation as well as in a myriad of other studies, does not always correspond with their scores on conventional, static tests. The results of the current study also indicate that, although all groups of children showed progression from session to session, there were also large individual differences between children, revealing individual differences in their potential for learning (e.g., Sternberg & Grigorenko, 2002).

Interestingly, when potential differences between the two groups of dynamically tested children categorised in the current study as gifted and average-ability are examined, a differential effect of training is not evident. Although the gifted children had significantly higher scores at each phase of the testing process, the progression lines of both groups demonstrated equivalent slopes. Although these findings contradict earlier research in which high IQ children were found to not only differ in their performance, but also have a broader zone of proximal development (e.g., Calero et al., 2011), they do suggest that dynamic testing could be applied successfully amongst children of all levels of intelligence. Our study found that the learning progress of gifted children was, to a large extent, more similar than different to that of average-ability children. One explanation as to why we could not find a difference in the breadth of the zone of proximal development could be that in previous research (Calero et al., 2001; Kanevsky, 2000) a higher cut-off score of cognitive functioning (than our use of the 90th centile) was used making the group of gifted children in previous studies more distinct. Another explanation might be found in a potential ceiling effect, although the most difficult analogy items required six transformations in order to solve them correctly. Moreover, in previous studies the same analogy items were solved by children of up to eight years old, and the authors of these studies do not mention a ceiling effect amongst their participants (e.g., Hosenfeld et al., 1997; Tunteler et al., 2008).

The second main aim of the current study was to investigate the association between test anxiety scores and progression in test performance after dynamic testing. Our findings suggested, in general, that test anxiety and improvement

in accuracy across test sessions were related. More importantly, we found that test anxiety was related to training benefits; children with higher levels of test anxiety showed significantly more gain in accuracy than their peers with lower levels. A possible explanation for this notion can be found in the literature. Meijer (2001) found, for example, that test anxiety stems from a lack of self-confidence. Related to this, Beckmann, Beckmann, and Elliott (2009) found that providing feedback to learners with low self-confidence can have a compensatory effect on performance, and help them achieve a level of performance approaching, or similar to, their peers with high self-confidence. In this respect, our findings mirror Beckmann and colleagues' (2009) findings. It seems plausible that a dynamic training intervention can also boost a child's self-confidence, although follow-up studies are needed to research this tentative conclusion. These findings supported, once more, the notion that testing children dynamically instead of statically could indeed lead to less biased test results (Sternberg & Grigorenko, 2009; Meijer, 1996, 2001).

In contrast to our expectations, we did not find differential training benefits amongst gifted and average-ability children with higher levels of test anxiety. This finding seems plausible in light of the fact that no differences were found in test anxiety scores, nor in progression after dynamic testing across the two ability groups. The finding that gifted and average-ability children's progression paths after being dynamically trained developed similarly, did not lend support to Zeidner and Shani-Zinovich's (2011) hypothesis. These findings do suggest, at the very least, that providing children, irrespective of their intellectual ability, with a dynamic training session weakens the relationship between test anxiety and performance in test situations. Although our results seem to suggest that dynamic testing also diminished test anxiety during the post-test, as also found by Bethge et al. (1982), this cannot be confirmed definitively. Two task-specific measurements of test anxiety would be required to investigate this issue more thoroughly – one prior and one after administration of the dynamic test.

The current study had some additional limitations. Firstly, it employed a short training session only, with no follow-up. Secondly, test anxiety scores were based on the children's self-reports. A question remains to what extent our findings can be generalised to children suffering from clinical levels of test anxiety. Thirdly, none of the children who participated in the current study were identified as strictly "gifted" prior to the study by means of full scale IQ testing. The Raven test, however, is widely considered to be a sound measure of general intelligence (or 'g'). Finally, aspects of gifted behaviour that are deemed important, such as creativity and task commitment (e.g., Kornilov, Tan, Elliott, Sternberg, & Grigorenko, 2012; Renzulli, 2002), were not assessed.

Finally, the study findings remind us that high cognitive potential does not automatically help such children to perform well in test situations. Therefore, we would recommend that children with high levels of test anxiety, should be tested dynamically, particularly in any situations where incapacitating stress is likely to impair their ability to demonstrate their true potential.