Dynamic testing and excellence: unfolding potential

Vogelaar, B.

Author: Vogelaar, B.
Title: Dynamic testing and excellence: unfolding potential
Issue Date: 2017-01-18

Dynamic Testing and Excellence

Unfolding potential

BART VOGELAAR

Cover design by: Roosmarijn Knijnenburg
Printed by: Ipskamp Printing B.V.
ISBN: 978-94-028-0420-1

Dynamic Testing and Excellence

Unfolding potential

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 18 januari 2017
klokke 15:00 uur

door

Bart Vogelaar
geboren te Krimpen aan den IJssel
in 1987

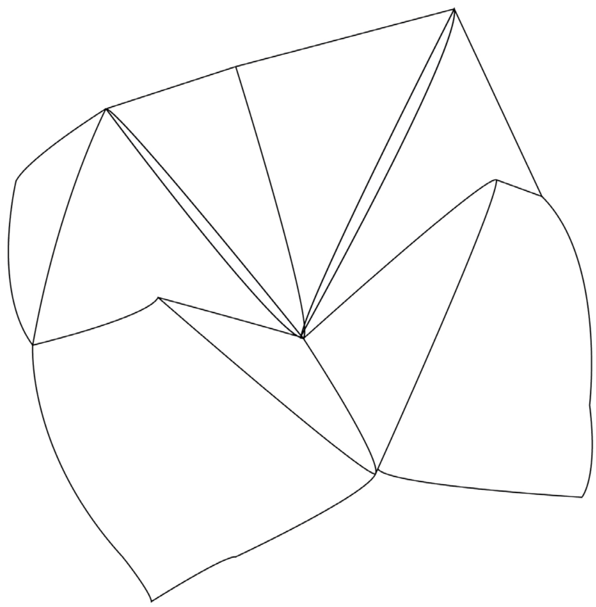
Promotor:
Prof.dr. W.C.M. Resing
Co-promotor:
Dr. A.J.M. Hoogeveen (Radboud Universiteit)
Promotiecommissie:
Prof.dr. E.A.M. Crone
Prof.dr. K.H. Wiedl (University of Osnabrück)
Prof.dr. M.G.P. Hessels (University of Geneva)
Dr. D. Zbainos (Harokopio University of Athens)
Dr. T. Bosma (Vrije Universiteit Amsterdam)

Contents

CHAPTER 1

General Introduction

1



In many countries, gifted and talented children have long been "forgotten" in education. Educational professionals have commonly held the view that these children would manage on their own, and would not benefit from additional help or extra attention (Persson, 2010). Recently, however, more attention is being paid to gifted learners in education, in general as well as in relation to their educational needs (Swanson, 2016; VanTassel-Baska & Stambauch, 2005). In order to be able to cater to gifted learners' educational needs, reliable and valid identification of gifted students is key. Although some researchers and practitioners have argued that static measures of intelligence are inadequate for these purposes (Lohman & Gambrell, 2012; Worrell & Erwin, 2011), shortened measures of intelligence are often used in identification of giftedness (Nisbett, 2009; Pierson, Kilmer, Rothlisberg, & McIntosh, 2012).

Conventional, static tests are, however, said to capture only part of cognitive functioning (e.g., Elliott, 2003; Lohman & Gambrell, 2012). Moreover, these tests are believed to measure previous learning experiences, which do not always correspond with children's cognitive potential (Sternberg & Grigorenko, 2002). Research has suggested that some children are disadvantaged when taking these tests, including those with a low socio economic status, a different cultural background, special needs (Robinson-Zañartu & Carlson, 2013), or test anxiety (Meijer, 1996). Moreover, static tests are primarily focused on testing outcomes, taking into account psychological processes involved in learning only indirectly (Jeltova, Birney, Fredine, Jarvin, Sternberg et al., 2007). This seems in contrast with the aim of gifted education, and education in general: unfolding and maximising talent and potential of each individual child (Nicpon & Pfeiffer, 2012). Gaining insight into how an individual (gifted) child learns, and what conditions are optimal for this child to demonstrate his or her potential therefore seem a necessity. This notion seems even more salient, as practitioners and educational professionals have voiced concerns that tailoring to the needs of gifted learners often proves difficult, which could ultimately result in loss of cognitive potential (Ryan & Coneybeare, 2013).

Different from static testing, which primarily measures previous learning, dynamic testing is focused on what children can learn in a short time-frame, by intertwining feedback and instruction in the testing process (Elliott, 2003). In that sense, it is considered to measure children's potential for learning (Elliott, Grigorenko, & Resing, 2010). As learning new skills is an important part of a dynamic test, this form of testing seems a promising tool for gaining insight into aspects and processes that play a role in learning (Resing, 2013). This thesis aimed to provide more insight into the processes and cognitive aspects playing a role in the learning process of children in general, and gifted children in particular.

Dynamic testing

Dynamic tests exist in many different forms, but they all have in common that they provide instruction, help, or feedback as part of the testing procedures, and, in doing so, are aimed at structured measuring of children's progression in learning (Sternberg & Grigorenko, 2002). This is one of the main objectives of dynamic testing, while providing insight into the type of help a child needs in order to learn is another key objective (Resing, 2013). Generally speaking, there are two dynamic testing formats: the cake, and the sandwich format (Sternberg & Grigorenko, 2001). As part of the cake format, a child is offered help, item by item, as soon as he or she experiences a substantial difficulty in solving a task. The sandwich format, also known as the pre-test-training-post-test design, is used in the current thesis.

Whereas some dynamic tests offer tailored, individual prompts, help, or feedback, other tests provide standardised instruction. In this thesis, children were provided with standardised prompts during the training phase based on graduated prompting principles. Graduated prompting has repeatedly shown to lead to progression in learning (Bosma & Resing, 2006; Ferrara, Brown, & Campione, 1986; Resing, 2013; Resing & Elliott, 2011). Graduated prompting refers to a form of training as a part of which children are provided with a prompt as soon as they experience a significant difficulty in solving a task (Campione & Brown, 1987; Resing & Elliott, 2011). In the studies that are part of this thesis, prompts were tailored to each test item, and provided hierarchically; ranging from very general metacognitive prompts, to task-specific cognitive prompts, to modelling (Resing, 2000). Providing these prompts in a hierarchic fashion enables measuring of different degrees of help each individual child needs in order to demonstrate learning.

Different measures are used as potential for learning measures, including the amount and the type of feedback received during training, performance after training, the change in performance from the pre-training to the post-training stage, and the transfer of learned skills (Elliott et al., 2010; Sternberg & Grigorenko, 2002). Examination of these different dynamic measures has revealed that there is great variability between, but also within, children in relation to their instructional needs (e.g., Bosma & Resing, 2006; Jeltova, Birney, Fredine, Jarvin, Sternberg et al., 2011), their level of improvement (e.g., Fabio, 2005; Resing & Elliott, 2011), and the degree to which they could transfer their learned skills (e.g., Resing, Bakker, Pronk, & Elliott, 2016; Tzuriel, 2007).

Dynamic testing of analogical reasoning

Many dynamic tests employ inductive reasoning tasks (e.g., Ferrara et al., 1986; Resing, 2013). Inductive reasoning has been noted for its central role in a variety of cognitive skills and processes (Csapó, 1997), such as general intelligence (Klauer & Phye, 2008), problem-solving (Richland & Burchinal, 2012), and acquisition and application of knowledge (Goswami, 2012). Inductive reasoning has been found to develop throughout childhood, during primary and secondary school (Csapó, 1997; Molnár, Greiff, & Csapó, 2013). In the studies part of this thesis, visuo-spatial analogical reasoning matrices were employed of the type A:B::C:?. These tasks were utilised, as opposed to more traditional scholastic tasks, for example in the domains of reading or maths, as these skills are often taught in several gradations with differences between schools being apparent. Measuring these skills would, therefore, require using a multitude of different dynamic tests. Analogical reasoning, a subtype of inductive reasoning, is said to play an important role in children's everyday learning (Richland, Morrison, & Holyoak, 2006). Moreover, research has shown that performance on analogical reasoning matrices, such as those used in the current thesis, is associated with scholastic achievement (Balboni, Naglieri, & Cubelli, 2010), as well as individual differences in IQ scores, as well as in fluid intelligence (Caropreso & White, 1994; Vendetti, Wu, & Holyoak, 2014).

Giftedness

Dynamic testing has originally been developed as a response to the fact that the cognitive abilities of children with special educational needs, or those with disadvantaged educational experiences, were not captured adequately by traditional testing instruments (Sternberg & Grigorenko, 2002). It is, therefore, not surprising that the vast majority of studies into dynamic testing have focused on such special groups of children. Over the past few decades, however, a few studies have been conducted in which dynamic testing was utilised to assess the cognitive abilities of gifted children. The focus of most of these studies was on identification of giftedness in ethnic and linguistically diverse populations for participation in gifted programmes (e.g., Lidz & Macrine, 2001; Matthews & Foster, 2005). Research further suggests that gifted children have a broader zone of proximal development, and demonstrate higher transfer rates (Calero, García-Martín, & Robles, 2011; Kanevsky, 1995, 2000). Moreover, Kanevsky (2000) concluded that the learning of gifted children was characterised by high levels of motivation, metacognition, self-regulation, and cognitive flexibility. Studies that systematically compare the cognitive abilities of gifted and typically developing children by means of dynamic testing are, however, scarce.

Although increasingly more attention is currently being paid to gifted children (see Dai, Swanson, & Cheng, 2011 for an overview), generalising research findings to practical solutions for identification and education of gifted children has revealed to be challenging (VanTassel-Baska, 2006). Researchers and practitioners alike have voiced their concerns in relation to the fact that there is no consensus on a definition of giftedness (Dai & Chen, 2013; Nicpon & Pfeiffer, 2011). Over the past century, there has been a shift in the general view that giftedness is not a unidimensional, but rather a multidimensional construct. While in one of the first definitions, giftedness was equalled by an IQ of at least 140 (Terman, 1925), already in the 1970s, although expanded on in more recent years (e.g., Renzulli, 2005; Renzulli & D'Souza, 2014; Renzulli & Sytsma, 2008), Renzulli (1978) noted that giftedness occurs in the interplay of above average cognitive capacities, creativity and task persistence.

At the beginning of the 21st century, definitions of giftedness started to emerge that also took into account interaction with the environment, and sociocultural content (Barab & Plucker, 2002). One of the most recent developments in the conceptualisation of giftedness is the model posited by Subotnik, Olzewski-Kubilius, and Worrell (2012) who define giftedness as

"performance that is clearly at the upper end of the distribution in a specific talent domain even relative to other high-functioning individuals in that domain. Further, giftedness can be viewed as developmental in that in the beginning stages, potential is the key variable; in later stages, achievement is the measure of giftedness; and in fully developed talents, eminence is the basis on which this label is granted" (p. 176).

As stated above, there is a gap between practice and research in relation to the identification and conceptualisation of giftedness (see e.g., Nicpon & Pfeiffer, 2011). In the USA, for instance, in most states giftedness is identified by an IQ test score only, and cut-off scores of at least the 90th or 95th percentile are common (McClain & Pfeiffer, 2012). The American National Association for Gifted Children (NAGC) considers individuals as gifted when they demonstrate outstanding reasoning and learning abilities or competence, operationalised as performance in – at least – the top 10% region, in one or more domains (NAGC, 2010). In addition, teacher nominations are frequently used to determine whether a child is gifted and should be placed in special settings for the gifted (Kornmann, Zettler, Kammerer, Gerjets, & Trautwein, 2015; Threlfall & Hargreaves, 2008). Participants in the studies in this thesis were selected randomly from primary schools in the western part of the Netherlands. Gifted children were oversampled, and were all enrolled in special settings for gifted and talented children in the western part of the Netherlands. The selection criteria for participating in the studies in this thesis consisted of parents' and teachers' nominations, in addition to their enrolment in these settings. In addition, the children participating in the studies described in Chapters 3, 4, and 5, in accordance with the NAGC, were found to have a percentile score of at least 90 of the Raven Standard Progressive Matrices Test (Raven, 1981). As described above, although various factors play a role in the conceptualisation of giftedness, in this thesis only cognitive factors are considered.

Developing Expertise Model

An alternative view on giftedness was posited by Sternberg (1999, 2001; Sternberg, Jarvin, & Grigorenko, 2011), who views giftedness as developing expertise. In his model of developing expertise, which links the measurement of potential for learning with giftedness, learning new abilities is similar to the development of expertise. Sternberg's model posits that five elements, metacognition, motivation, knowledge, thinking, and learning itself, play a role in becoming an expert. These elements are interactive, influencing each other both directly and indirectly. Through practice, a novice can become an expert in a specific learning context. Sternberg (2001) further proposed that giftedness equals an exceptional ability to develop expertise, within a zone of proximal development, on the basis of existing or developing abilities and skills. In his view, gifted children have greater potential for developing expertise, they develop expertise at a faster rate, to higher levels, or to qualitatively different levels than non-gifted children.

In this thesis, the relationship between a number of the factors described in the Developing Expertise Model and dynamic testing outcomes were examined; specifically two aspects of executive functioning, cognitive flexibility and metacognition, as well as the ability to generalise knowledge and skills (transfer). Moreover, it was investigated whether test anxiety would be related to these test outcomes.

Executive functioning

Executive functioning plays a central role in developing expertise, as posited in Sternberg's Developing Expertise Model. In line with his model, gifted children are often said to have an executive functioning advantage (Arffa, 2007). Executive functioning is an umbrella term used to refer to a number of complex cognitive processes enabling conscious control of thought and action that are critical to purposeful, goal-directed behavior (Monette, Bigras, & Guay, 2011). Executive functioning is deemed important when learning new skills (e.g., Diamond, 2013) and has been found to be a predictor of academic success (Viterbori, Usai, Traverso, & De Franchis, 2015).

Measuring executive functioning is often considered as challenging (e.g., Miyake et al., 2000; Viterbori et al., 2015). In general, two types of instruments are used to obtain measures of executive functioning: (self and informant) rating scales and performance-based tasks. A difficulty of measuring executive functioning is that most performance-based tasks have originally been developed for adults (Isquith, Crawford, Andrews Espy, & Gioia, 2005). Using these instruments for children has several implications, particularly when taking into account the developmental nature of executive functions. Studies have indicated that executive functions develop throughout childhood until late adolescence (Gathercole, Pickering, Ambridge, & Wearing, 2004; Huizinga, Dolan, & Van der Molen, 2006). In addition, some researchers have also noted that rating scales do not always fully capture children's executive functioning (e.g., Sadeh, Burns, & Sullivan, 2012). Due to these reasons, researchers recommend the use of various instruments when assessing executive functions (Toplak, West, & Stanovich, 2013). In the present thesis, the influence of cognitive flexibility, the ability to be flexible in adjusting thinking to meet changing demands (Diamond, 2013), measured by means of a performance-based task, and metacognition in general, measured by a teacher rating scale, was examined in relation to static versus dynamic measures of analogy problem-solving.

Test anxiety

In addition to elements that facilitate learning and the development of expertise, there are also factors that may hinder learning. One of these factors is test anxiety. The adverse effects of test anxiety on cognitive performance have been well-documented, ranging from, for instance, scholastic achievement (e.g., Segool, Carlson, Goforth, Von der Embse, & Barterian, 2013), to intelligence testing (e.g., Hopko, Crittendon, Grant, & Wilson, 2005; Meijer, 1996, 2001). Test anxiety has even been described as one of the causes of underperformance of various learners, including gifted children (e.g., Reis & McCoach, 2010). It is estimated that 10 to 40 per cent of all students have experienced clinical levels of test anxiety (Segool et al., 2013). It has been stated in the literature that in the gifted population, test anxiety is less prevalent; these children are assumed to have higher intellectual coping resources that may lead to them coping better in stressful academic situations (Zeidner & Shani-Zinovich, 2011).

Test anxiety is commonly measured by means of self or informant report scales (e.g., Wren & Benson, 2004). Some research, using self-report measures, has suggested that testing children dynamically rather than statically resulted in lower test anxiety levels in primary school children (Bethge, Carlson, & Wiedl, 1982). These researchers found that amongst third grade children, test anxiety seems to be diminished when children's ability to solve analogies was assessed dynamically. Research into the relationship between test anxiety and dynamic testing in gifted children, however, is scarce (for studies on test anxiety and dynamic testing, in general, see e.g., Bethge et al., 1982; Meijer, 1996; 2001). In this thesis, it was examined whether test anxiety, as measured by a self-report questionnaire, has a differential influence on static and dynamic measures of analogy problem-solving.

Transfer

Transfer is the ability to apply and adapt knowledge to a new context, and is an important goal of education (Day & Goldstone, 2012). In spite of the fact that transfer has been examined for more than a century (Engle, 2012), eliciting transfer of learning has proven to be difficult (Day & Goldstone, 2012). It is assumed that two specific factors are important in the effectiveness of transfer: the content and the context. The content refers to the actual content being transferred (Barnett & Ceci, 2002), whereas the context is used to denote the different domains from and to which transfer takes place (Klahr & Chen, 2011). Different types of transfer have been proposed, which are based on the extent to which the base and target problem share similarities (Barnett & Ceci, 2002). Often, transfer is classified in terms of near versus far transfer (Mestre, 2005), and surface versus deep transfer (Forbus, Gentner, & Law, 1995).

Although the underlying processes of transfer are still not fully understood, research suggests that successful transfer is associated with the extent to which an individual child mastered the task to be transferred (Siegler, 2006). Expertise, and a deep understanding of the task at hand seems to be required in order for deep transfer to be successful (Barnett & Ceci, 2002). Some studies suggested that gifted children outperform their average-ability peers with regard to the extent to which they demonstrate successful transfer (e.g., Klavir & Gorodetksy, 2001), but other studies have supported this conclusion only partially; on near transfer tasks gifted and average-ability children show similar rates of transfer (Carr, Alexander, & Schwanenflugel, 1996), while in other studies on far transfer tasks gifted children demonstrated superior transfer rates (Geake, 2008; Kanevsky, 2000). In the current thesis, transfer was investigated by utilising a 'reversal' procedure, in the form of an analogy construction task (e.g., Bosma & Resing, 2006; Harpaz-Itay, Kaniel, & Ben-Amran, 2006). The potential role of training on transfer success and effectiveness was investigated, as well as the roles of giftedness, and mastery of analogy problem-solving.

Outline of this thesis

The current thesis utilised dynamic testing principles to investigate potential differences between gifted and average-ability children in relation to their potential for learning, instructional needs, and their ability to transfer learned skills. A number of factors described in Sternberg's (1999; 2001; Sternberg et al., 2011)

Developing Expertise Model that possibly influence dynamic testing outcomes were examined, with a specific focus on executive functioning. Children's progressions in analogy problem-solving and analogy construction were considered, taking into account the roles that age, ability, executive functioning, and test anxiety played.

In Chapter 1, the studies that are part of this thesis were introduced, and a theoretical background for these studies was provided. In Chapter 2, dynamic testing principles were employed to examine potential differences in progression of analogy problem-solving of gifted and average-ability children of 5-8 years old. Taking into account age, it was investigated whether gifted and average-ability children demonstrated differential progression in analogy problem-solving, benefitted differentially from a dynamic training procedure, and showed differential instructional needs.

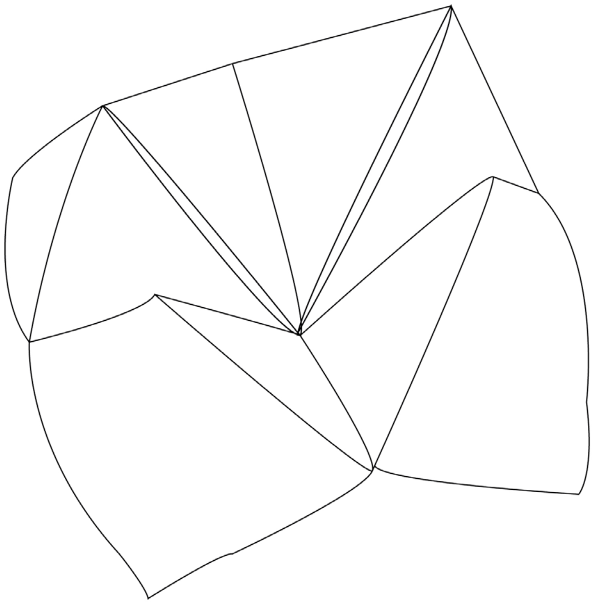
In Chapters 3 and 4, we further examined 7 and 8 year old gifted and average-ability children's progression in analogy problem-solving, using Linear Mixed Modeling analysis with a multilevel approach. In both chapters, the growth trajectories of children in the various conditions were investigated. In Chapter 3, the potential role of test anxiety was examined in relation to static and dynamic progression in analogy problem-solving. In Chapter 4, analogy problem-solving was examined in relation to two aspects of executive functioning: cognitive flexibility, as measured by a performance-based task, and metacognition in general, as measured by a teacher rating scale. Potential differences in instructional needs of gifted and average-ability children were further explored in this chapter.

In Chapter 5, the main focus was on children's transfer of analogy problem-solving, which was examined by means of an analogy construction task. Potential differences in transfer performance and the degree of transfer of analogy problem-solving of 9 to 10 year old gifted and average-ability children were examined. It was investigated whether ability, training and analogy problem-solving performance were associated with rates of transfer. In Chapter 6, the results of the studies part of this thesis were discussed, as well as the implications of the key findings for research, assessment and education, in particular in relation to gifted children.

CHAPTER 2

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

2



Bart Vogelaar
Wilma C. M. Resing

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

Abstract

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

2.1 Introduction

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

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

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

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

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

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

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

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

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

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

Our second cluster of research questions concerned potential differences in instructional needs amongst children identified as gifted and average-ability children, taking into account age. We hypothesized that the children identified as gifted would need less help to solve the analogies than their average-ability age-mates, and that they would, more specifically, need less cognitive help, because general, metacognitive help would, presumably, suffice in order for them to accurately solve the problems given. This hypothesis builds uon Kanevsky's (1994) findings that gifted children were more responsive to feedback and that they were assumed to have an advantage in self-regulation (see also Calero et al., 2007; Zimmerman, 1989). We further explored whether age would play a role in the instructional needs of all dynamically tested children and expected that considering that older children were found to be better at solving analogies (Csapó, 1997; Hosenfeld, Van den Boom et al., 1997), they would need less help than the younger children, regardless of their ability.

2.2. Method

Participants

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

Design

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

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

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

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

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

Materials

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

Analogy tasks: Tasks and Dynamic Training Procedure. In this study, a series of visuospatial analogy tasks had to be solved, assumed to measure inductive reasoning (Barnett & Ceci, 2002). During the unguided practice sessions, children were provided with series of 20 equivalent, parallel items, composed of geometric analogies of varying difficulty of the type A:B::C:?. All series had different items of comparable item difficulties. The test sessions were equivalent in terms of item difficulty variation, and the order in which the items were presented, but differed in the sense that each test session was composed of new analogy items. These items were a selection of a test battery originally created by Hosenfeld, Van den Boom et al. (1997) and adapted by Tunteler et al. (2008; see Figure 1 for an example). In the construction of all items, six basic geometrical shapes were used: squares, triangles, hexagons, pentagons, circles, and ovals. Each analogy was constructed by means of five possible transformations: changing position, adding or subtracting an element, changing size, halving, and doubling. The test was administered as a paper-and-pencil test, and the children were asked to draw the correct answers. Because the children were asked to construct and draw their answers themselves, a test session could not have more than 20 items because of time constraints. Test Session 1 (the pre-test) had a high level of internal consistency, as determined by a Cronbach's alpha of .94 for the accuracy scores

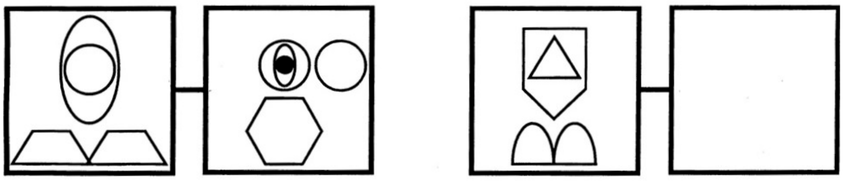


Figure 1. Example of a difficult analogy item.

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

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

2.3. Results

Descriptive data

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

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

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

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

Changes over time in progression of accuracy

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

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

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

To examine our first cluster of hypotheses, a repeated measures (RM) analysis of variance (ANOVA) with one within-factor Session (Session 1, Session 2, Session 3) and three between-factors Treatment (unguided control vs. dynamic training), Ability Category (identified as gifted vs. average-ability), and Age (5-6 vs. 7-8 years) was conducted with the number of accurately solved analogy items at the three sessions as the dependent variable. The results showed, as expected, significant between effects for Ability Category and Age, $F(1, 258) = 33.23$, $p < .001$, $\eta_p^2 = .12$ and $F(2, 258) = 107.94$, $p < .001$, $\eta_p^2 = .30$, respectively, and a significant main session effect, $F(2, 516) = 151.72$, $p < .001$, $\eta_p^2 = .37$. Contrast analysis and visual inspection revealed significant progressions from Session 1 to Session 2 and Session 2 to Session 3, $F(1, 258) = 214.07$, $p < .001$, $\eta_p^2 = .45$, and $F(1, 258) = 21.48$, $p < .001$, $\eta_p^2 = .08$, respectively, indicating that all groups of children progressed significantly in their accuracy to solve analogies from one session to the next. The RM analysis, as expected, further showed a significant Session x Treatment interaction, $F(2, 516) = 12.62$, $p < .001$, $\eta_p^2 = .05$. Contrast analysis showed, as expected, only a significant interaction from Session 2 to Session 3, $F(1, 258)=21.29$, $p<.001$, $\eta_p^2=.08$, indicating that the dynamically trained groups of children after Session 2 outperformed the groups of children that had unguided practice experiences (see also Figure 2). However, in contrast with our expectations, no significant interactions between Ability Category and Treatment, Ability Category and Session, or Ability Category and Treatment and Session were revealed. These findings indicate that children who were categorized as gifted did have higher scores in analogical reasoning in general but, regardless of whether they were trained or not, showed parallel progression paths when compared with the group of children who were not categorized as gifted. A significant Session x Age effect, $F(2, 516) = 5.25$, $p = .006$, $\eta_p^2 = .02$, followed by contrast analysis (only significant from Session 1 to Session 2, $F(1, 258) = 4.37$, $p < .04$, $\eta_p^2 = .02$), showed that the age groups only differed in the extent to which they progressed from Session 1 to Session 2, but not from Session 2 to Session 3.

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

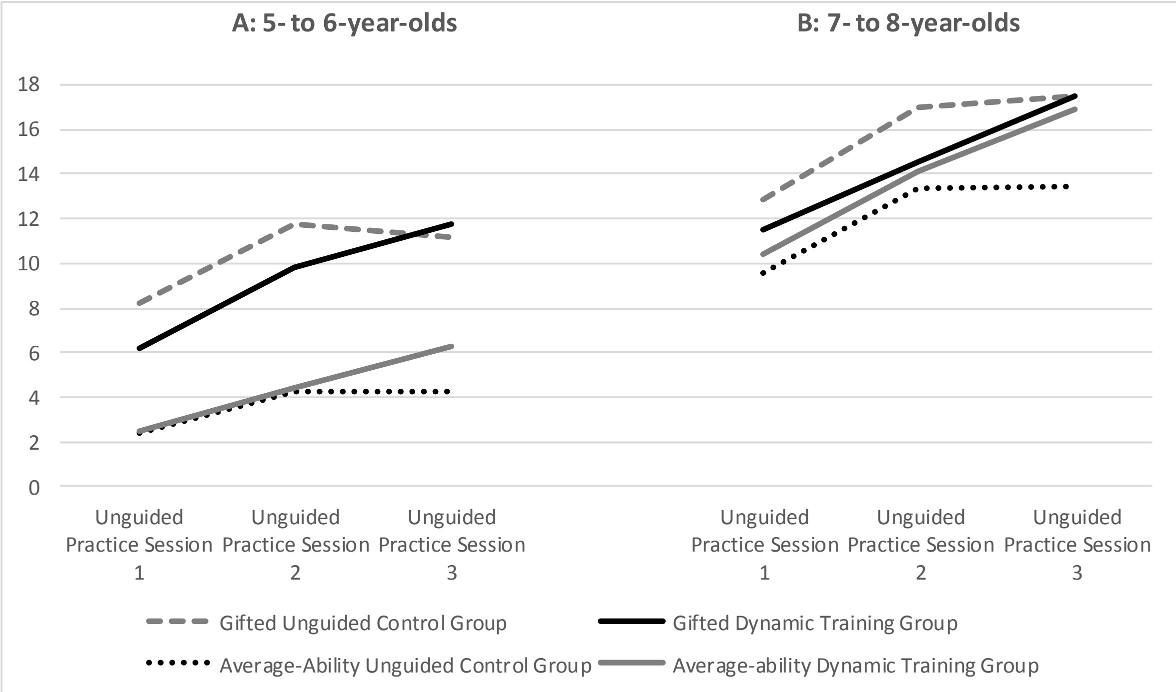


Figure 2. Mean Scores of the Number of Items Correct on session 1,2, and 3 per subgroup, divided by age in two separate parts for clarity reasons: 5- to 6-year-olds **(A)** and 7- to 8-year-olds **(B)**.

Need for instruction

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

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

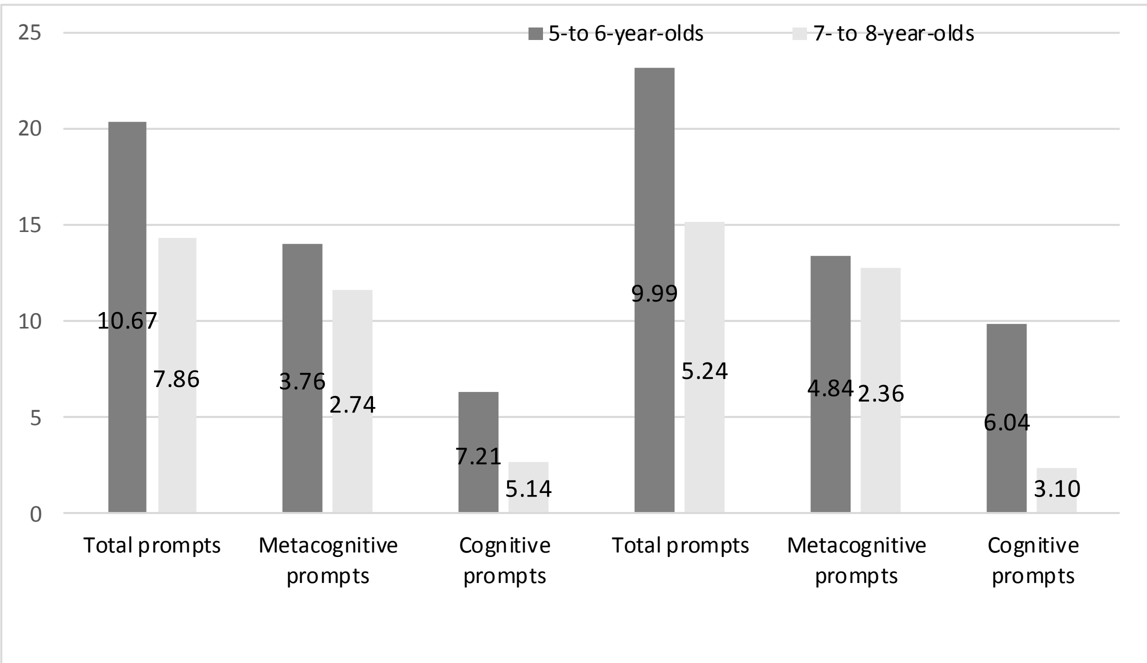


Figure 3. Mean scores and standard deviations (as shown in the individual bars) of the total number of prompts, metacognitive, and cognitive prompts received during training by talent and age.

2.4. Discussion

This study sought to examine whether two groups of children, a group identified as gifted through qualitative judgments by their teachers and parents, and a group of average-ability children, profited differently from unguided practice, and the dynamic testing intervention. Our aim was to obtain more insight into differences in performance of visuospatial analogical reasoning tasks, and differences in instructional needs, taking into account different age groups. Unlike several studies in the field in which significant differences between gifted and non-gifted children have been found in their performance on a dynamic test (e.g., Calero et al., 2011; Kanevsky, 1990, 1992, 2000; Lidz & Macrine, 2001), and their progression after unguided practice (e.g., Steiner, 2006) were revealed, the children categorized as gifted in this study showed similar, rather than different, progression paths to their average-ability age-mates, while starting at a higher initial ability point than the average-ability and outperforming their average-ability age-mates at each session, regardless of training and age.

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

The dynamic testing intervention, however, indeed seemed to lead to additional progression in analogical reasoning from the second session, before training, to the third session, after training, for all groups of children that seemingly could not be explained by practice alone, confirming previous research into graduated prompting techniques (e.g., Resing, 2000; Stevenson et al., 2013). The step-by-step tailored prompts seemed to provide the children with the tools they needed to progress beyond their accuracy scores before training. Looking more closely into the progression from the second session to the third, after training, in line with our expectations, it was found that both the dynamically trained children categorized as gifted, and the children categorized as average-ability benefitted from the dynamic testing intervention. It must be noted, however, that, in contrast with our expectations, and findings from earlier research (e.g., Calero et al., 2011), the children categorized as gifted did not benefit significantly more. Of course, it must be taken into consideration that the group of children, categorized in this study as average-ability, may have included children that, in fact, belonged to the gifted group but were not identified as such, for example, because of average school results. It is well-documented that children, and in particular gifted children, do not always live up to their potential for excellent performance, potentially as a result of character traits, motivation, internal mediators such as fear of failure, or incorrect usage of strategies (e.g., Reis & McCoach, 2000). It remains as yet unclear, however, whether this might have influenced our research findings.

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

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

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

Moreover, our findings regarding instructional needs showed that, irrespective of age, the gifted and average-ability children had similar instructional needs, which was in contrast with our expectations and findings from previous research (Calero et al., 2007; Kanevsky, 1990, 1994). Considering that all children who were categorized as gifted in this study attended gifted or talented education, this finding does hold important implications for gifted and talented education. Although it is generally assumed that gifted children manage their own classroom learning (De Boer et al., 2013), because they are said to be self-regulated learners and self-starters (e.g., Azevedo & Hadwin, 2005; Risemberg & Zimmerman, 1992), it seems that this does not necessarily mean that all gifted children have a need for less instruction. This research finding underlines the importance of using instructional and differentiation techniques in gifted and talented education, tailored to individuals' instructional and more general educational needs, for instance, by means of adaptive instruction. This is a type of instruction that aims to increase individual potential through performance demands appropriate for the individual (Heller, 1999). Considering that all the gifted children in this study were enrolled in either talented or gifted education, a type of education that in general aims to make as much use of high potential as possible (Dai & Chen, 2013), and endeavors to enable independent learning (Heller, 1999), it is surprising that their analogical reasoning progression in this study was not characterized by more independent learning. Future research could investigate this more closely, investigating whether the type of education influences the extent to which a child portrays independent learning, in the hopes of tailoring these types of education even more to the specific needs of talented and gifted children to achieve the best possible fit.

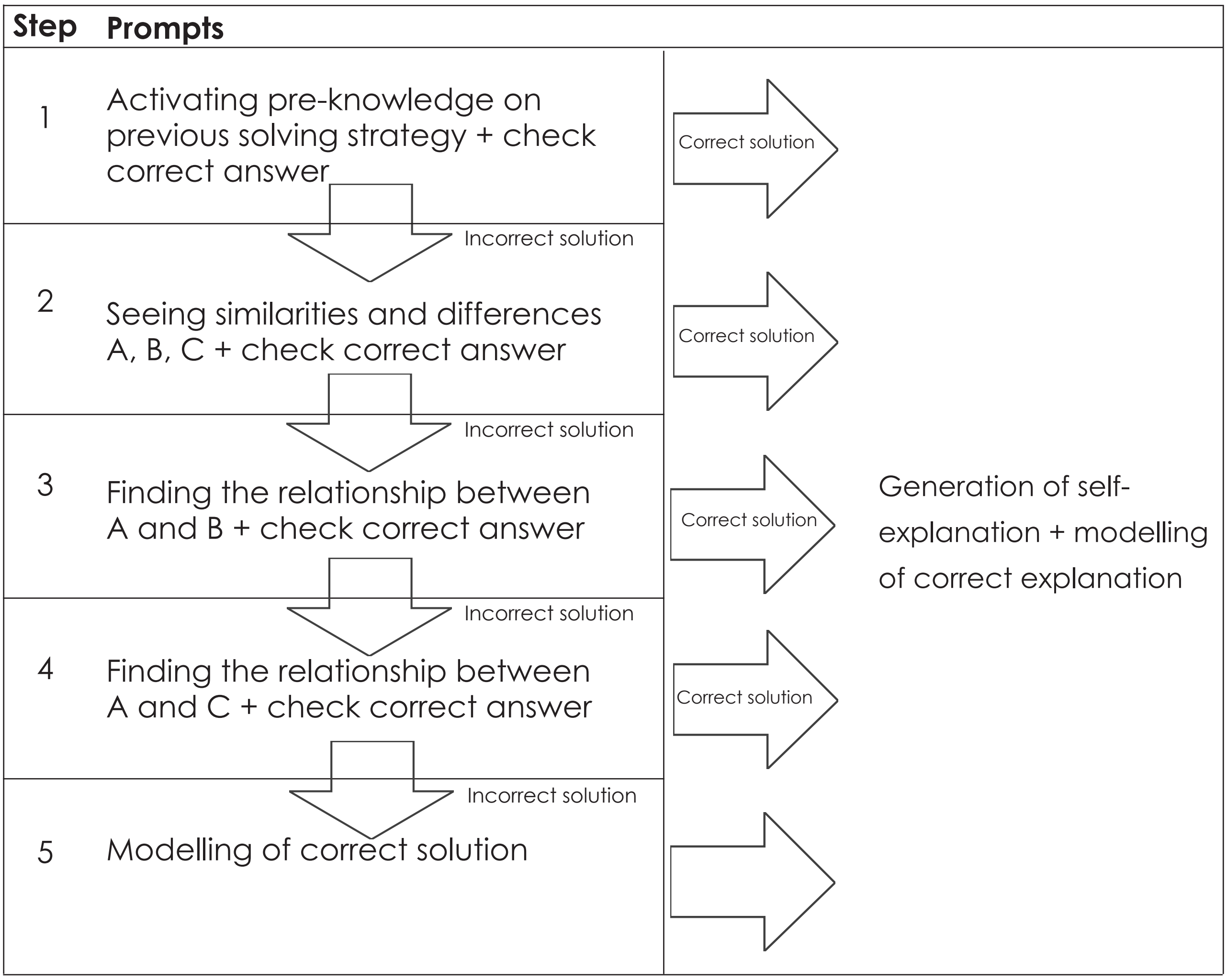
In addition to the limitations mentioned in the preceding text, our study had some other limitations. Our study looked into the learning progression of 5-to-8-year-old children. Because there were no norm scores available for 5-year-old children, we used the raw scores of the Raven instead of percentile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

The fact that the gifted children showed progression paths and instructional needs similar to the average-ability children, with variability in progression as well as instructional needs just like the average-ability children, and the fact that gifted children were, in general, found to have higher accuracy scores, ultimately suggests that dynamic testing can be used to measure the potential for learning of all children, including children with higher intelligence. The question that still needs answering is whether, and if so, to what extent, gifted, talented, and non-gifted children really differ qualitatively regarding their learning characteristics and processes (Dai & Chen, 2013). Our research results underline the importance and usefulness of combining microgenetic research results with dynamic testing procedures and gained more insight into potential differences in analogical reasoning development of young gifted and average-ability children. Hopefully, future research employing microgenetic techniques in combination with dynamic testing procedures could shed more light on this question. Important and promising areas to research in more detail employing these techniques would be strategy use and transfer because these are areas in which gifted children are assumed to differ significantly from non-gifted children in performance (e.g., Kanevsky, 1990; Scruggs & Mastropieri, 1988). Combining these topics in research might lead to a more detailed insight into the learning processes and educational needs of talented and gifted children, which would enhance our understanding of the underlying concepts involved. This, in turn, would greatly inform educational practice of these special groups.

Note

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

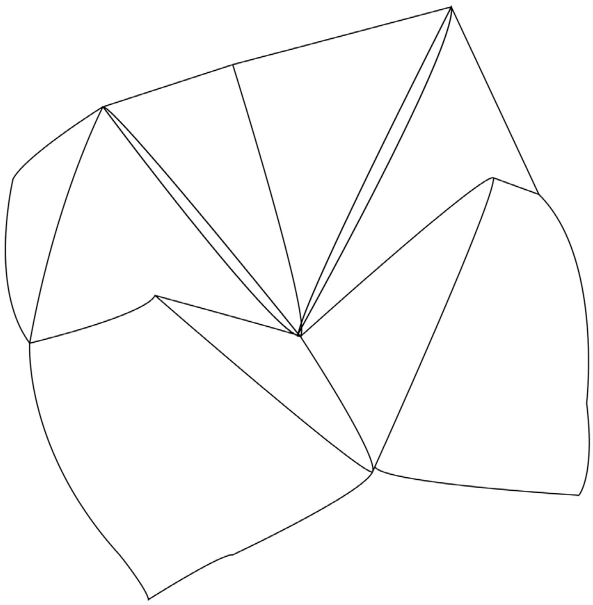
Appendix. Schematic Overview of the Graduated Prompts Training Protocol



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 -.2 between text 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) Average-ability (31)	X	X	X	Dynamic training	X	X
Unguided practice (N=60)	Gifted (23) Average-ability (27)	X	X	X	Dot-to-dot control task	X	X

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 α = .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.

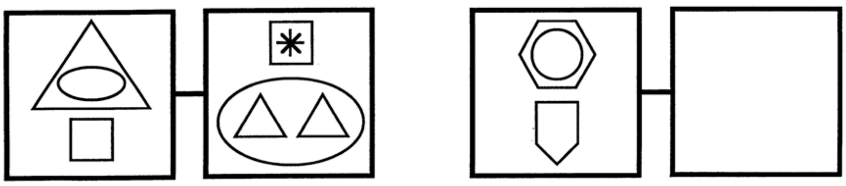


Figure 1. Example of a difficult analogy item.

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 α = .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.

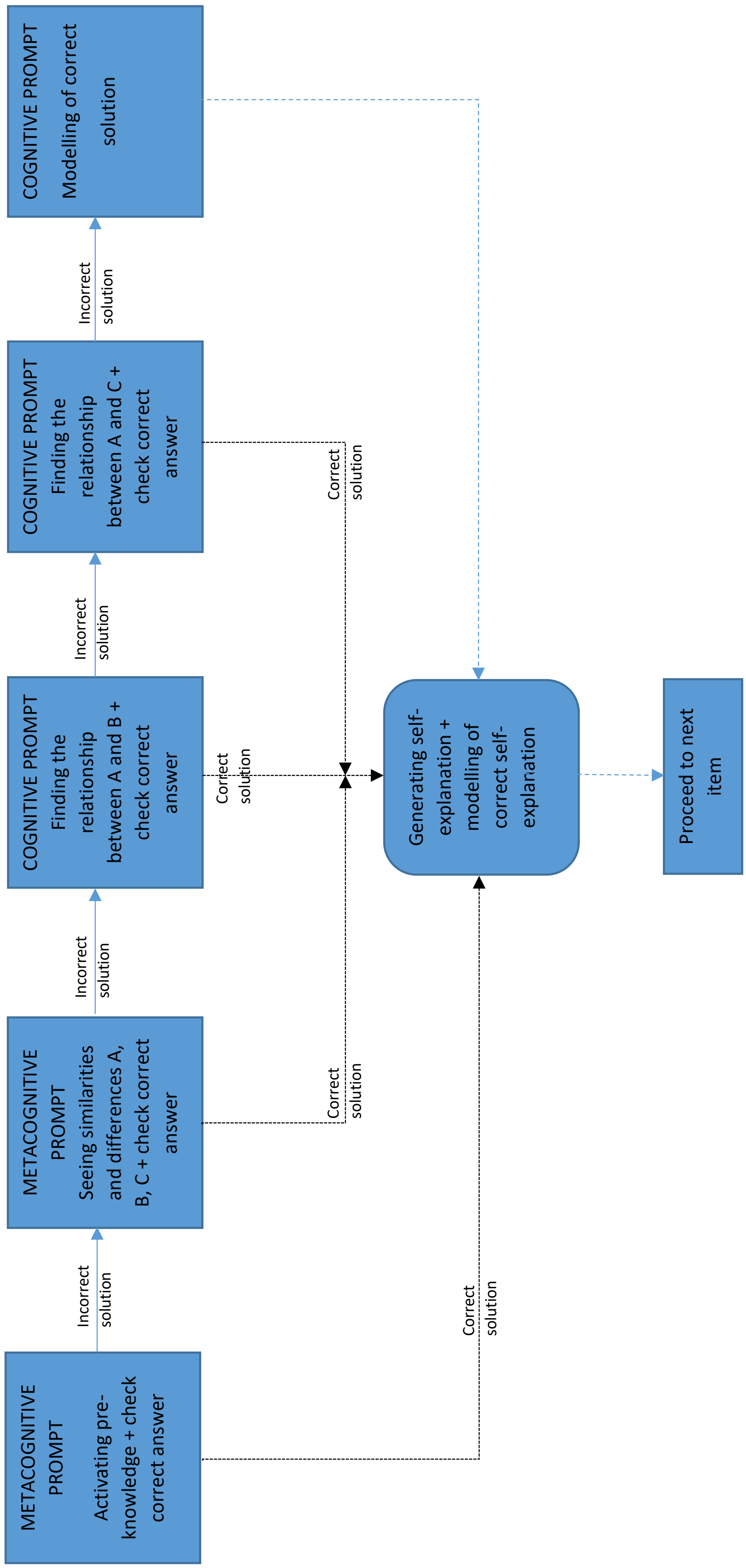


Figure 2. Flowchart of the graduated prompts training protocol.

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).

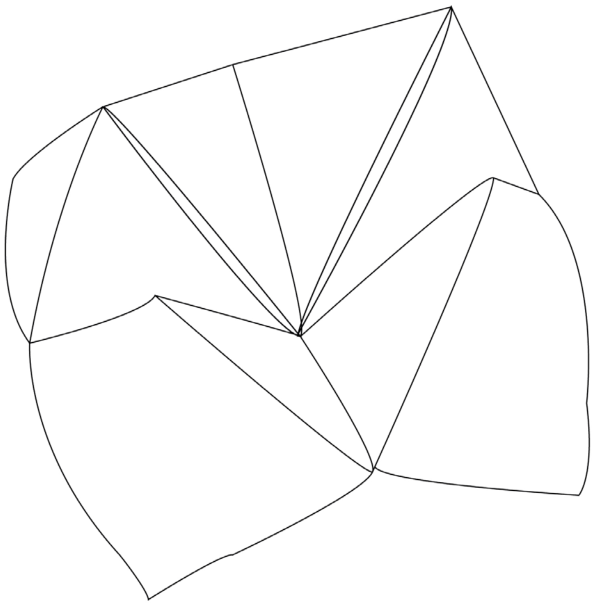
In contrast to our expectations, we did not find differential training benefits amongst gifted and average-ability children with higher levels of test anxiety. This finding seems plausible in light of the fact that no differences were found in test anxiety scores, nor in progression after dynamic testing 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.

CHAPTER 4

Dynamic testing of gifted and average-ability children's analogy problem-solving: Does executive functioning play a role?



4

Bart Vogelaar
Merel Bakker
Lianne Hoogeveen
Wilma C. M. Resing

Vogelaar, B., Bakker, M., Hoogeveen, L., & Resing, W. C. M. (submitted). Dynamic testing of gifted and average-ability children's analogy problem-solving: Does executive functioning play a role? *[Psychology in the Schools]*

Abstract

In this study, dynamic testing principles were applied to examine progression of analogy problem-solving, the roles that cognitive flexibility and metacognition play in children's progression as well as training benefits, and instructional needs of 7-8 year old gifted and average-ability children. Utilizing a pre-test-training-post-test control group design, participants were split in four subgroups: gifted dynamic testing (n=22), gifted unguided practice (n=23), average-ability dynamic testing (n=31) and average-ability unguided practice (n=37). Results revealed that dynamic testing led to more advanced progression than unguided practice, and that gifted and average-ability children showed equivalent progression lines and instructional needs. For children in both ability categories, cognitive flexibility was not found to be related to progression in analogy problem-solving or training benefits. In addition, metacognition was revealed to be associated with training benefits. Implications for educational practice were provided in the discussion.

4.1. Introduction

It has been proposed that cognitive abilities play an important role in children's school performance. Both intelligence (Balboni, Naglieri, & Cubelli, 2010; Roth, Becker, Romeyke, Schäfer, Domnick et al., 2015), and executive functions (e.g., Blair & Diamond, 2008; Monette, Bigras, & Guay, 2011; Viterbori, Usai, Traverso, & De Franchis, 2015) have been shown to predict school success. When a child is considered to be gifted in an educational context, this is often based on the results of an assessment procedure, including conventional, static testing of intelligence, or school aptitude (Kline, 2001). These tests, however, have been shown not to be advantageous for all children, and do not unveil information about psychological processes involved in learning (e.g., Grigorenko, 2009). As conventional tests, for a large part, rely on past learning experiences (Elliott, Grigorenko, & Resing, 2010), children who have had less than favorable learning experiences, have been documented to underperform on these tests (Robinson-Zañartu & Carlson, 2013). Dynamic tests, in contrast, are much more focused on a child's potential for learning, rather than on past learning experiences (Sternberg & Grigorenko, 2002). As in these tests, feedback and/or instruction are integrated into the testing procedure (Elliott, 2003), they allow for examining to what extent children show improvement in performance after an intervention, and whether other cognitive factors, such as executive functions, play a role in learning. In the current study, dynamic testing principles were applied to investigate to what extent two aspects of executive functioning, cognitive flexibility and metacognition, would be related to static or dynamic progression in analogy problem-solving of gifted and average-ability children.

Dynamic testing

Rather than measuring the knowledge or skills a child has already mastered, dynamic testing focuses on what a child would achieve in a short time-frame, and this assessment procedure is therefore expected to provide a more complete picture of a child's potential for learning (Elliott, 2003). The pre-test-training-post-test design (Sternberg & Grigorenko, 2002) is a frequently used application of dynamic testing that allows for structured measuring of a child's learning progression. The graduated prompts technique (e.g., Campione & Brown, 1987) has been used successfully as a training intervention in combination with said design. In this training approach, children are provided with structured prompts each time they make a mistake in problem solving. In the current study, prompts were tailored to each individual problem to be solved, and became more specific gradually, ranging from metacognitive to cognitive prompts and modelling (Resing & Elliott, 2011).

Similar to static test scores, dynamic testing outcomes have shown that there are many individual differences between children; both in terms of the instruction they require in order to show learning progression, as well as in terms of the level of progression they show after training (e.g., Resing, 2013, Sternberg & Grigorenko, 2002). Dynamic testing of children who have strong cognitive capacities, nevertheless, seems an area researched less intensively. Previous research indicates that gifted children not only have a cognitive advantage, but, more specifically, have a more extensive zone of proximal development, learn new skills faster, and are better at generalizing newly acquired knowledge (Calero, García-Martín, & Robles, 2011; Kanevsky, 2000). The potential role of executive functioning in dynamic testing of this group of children has, however, not yet been examined abundantly.

Dynamic tests frequently utilize inductive reasoning tasks (e.g., Ferrara, Brown, & Campione, 1986; Resing, 2000). Inductive reasoning is believed to play a central role in intelligence (Klauer & Phye, 2008), and is said to be of crucial importance with regard to acquiring and applying knowledge (Goswami, 2012) and solving problems (Richland & Burchinal, 2012).

Executive functioning

The graduated prompts technique employed in the current study included prompts activating different aspects of executive functioning, for example in relation to self-regulation and monitoring of the problem-solving process. Executive functions comprise a number of complex cognitive processes enabling conscious control of thought and action (Monette et al., 2011) that are critical to purposeful, goal-directed behavior (Arffa, 2007). They are seen as the cognitive component of self-regulation (Calkins & Marcovitch, 2010). Metacognition, a specific aspect of executive functioning, is usually described as consisting of self-reflective cognitive processes (Schneider, 2010), divided into two dimensions: knowledge, and regulation of cognitive activity (Moses & Baird, 1999), and is asserted to play an important role in developing new expertise (e.g., Sternberg, 1998).

In addition, it has been argued that flexibility in applying newly learned skills and knowledge can be seen as an important aspect of intellectual and cognitive functioning (e.g., Resing, 2013). Cognitive flexibility is said to include the ability to change perspectives spatially, or interpersonally, and being sufficiently flexible to adjust thinking to changing demands. Further, it is seen as a key component of the ability to think outside the box, and shares many characteristics with creativity, task and set switching (Diamond, 2013).

Executive functioning has been found to be related to cognition (e.g., Ardila, Pineda & Rosselli, 2000). Studies investigating the relationship of executive functioning in a dynamic testing context, in particular with gifted children, however, are few, with most studies focusing on the role of working memory (e.g. Resing, Xenidou-Dervou, Steijn, & Elliott, 2012; Stevenson, Bergwerff, Heiser, & Resing, 2014; Stevenson, Heiser, & Resing, 2013; Swanson, 2006, 2010, 2011).

The current study

The current study utilized a dynamic test for analogical problem solving, a subtype of inductive reasoning, employing graduated prompts techniques. As studies have shown that analogical reasoning develops greatly in 7-8 year old children (e.g., Richland, Morrison, & Holyoak, 2006; Tunteler & Resing, 2007), children of this age group participated in this study. Our main research aim was to provide more insight into the potential benefits of dynamic testing of gifted children. More specifically, we focused on the roles that ability, cognitive flexibility and metacognition play in repeatedly measured static versus dynamic progression in solving analogies.

Our first cluster of research questions addressed children's progression in solving analogies from pre-test to post-test. Based on previous research into progression of unprompted solving of analogy problems amongst young children (e.g. Tunteler & Resing, 2007; Tunteler, Pronk, & Resing, 2008), we expected a significant main effect of time. We hypothesized (1a) that both unguided practice, and dynamic testing would lead to progression in solving analogies from session to session. More importantly, we expected a significant interaction of time x condition, hypothesizing (1b) that children in the dynamic testing condition would show more progression from pre-test, before training, to post-test, after training (e.g., Resing & Elliott, 2011; Stevenson et al., 2013). As our study focused on potential differences between gifted and average-ability children in relation to their progression, we expected a significant interaction between time and ability. Gifted children were reported to have a more extensive zone of proximal development (e.g., Calero et al., 2011; Kanevsky, 2000), therefore we hypothesized (1c) that gifted children would show more progression after unguided practice experiences than their average-ability peers. We also expected a significant interaction of time x condition x ability, indicating that gifted children would show more progression after training than their average-ability peers (1d).

Our second cluster of research questions concerned the association between executive functioning and children's progression from pre-test to post-test. We expected a significant interaction between time and cognitive flexibility. Considering that flexibility in applying skills and knowledge is suggested to be important for learning and applying new knowledge (e.g., Resing, 2013), we hypothesized (2a) that children with higher levels of cognitive flexibility would show more progression in solving analogies than their peers with lower levels of cognitive flexibility. We also expected an interaction between time, condition, and cognitive flexibility, (2b) hypothesizing that children with higher levels of cognitive flexibility would benefit more from dynamic training than those with lower levels. Furthermore, a significant interaction between time, condition, ability and cognitive flexibility was expected. Building on empirical studies in which high-ability children were found to have an advantage in executive functioning (e.g., Arffa, 2007), we hypothesized (2c) that the progression paths of gifted children with higher levels of cognitive flexibility would be steeper than those of their average-ability peers with similar levels of cognitive flexibility.

Moreover, as self-regulating, metacognitive skills were found to play a significant role in learning (e.g., Campione, Brown, & Ferrara, 1982; Sternberg, 1998), we expected an interaction between time and metacognition, hypothesizing (3a) that children with higher levels of metacognition would show more progression in solving analogies than their peers with lower levels of metacognition. We also expected a significant interaction between time, metacognition and condition, and hypothesized (3b) that children with higher levels of metacognition would benefit more from training than their age-mates with lower levels of metacognition. Finally, a significant interaction was expected between time, condition, ability and metacognition. Taking into account that high-ability children were found to have an advantage in self-regulation (e.g., Calero, García-Martín, Jiménez, Kazén, & Araque, 2007), we hypothesized (3c) that the progression paths after training of the gifted children who have higher levels of metacognition would be steeper than their average-ability peers with similar levels of metacognition.

Our last research question focused on more closely to what extent gifted and average-ability children have different instructional needs, as measured by the number and the type of prompts required during training. As high-ability children were found to be more responsive to feedback (Kanevsky & Geake, 2004), and were found to have an advantage in self-regulation (e.g., Calero et al., 2007), we expected that gifted children's instructional needs during dynamic training would be significantly different from their average-ability peers. We hypothesized that gifted children would (4a) need both less metacognitive and (4b) less cognitive prompts than their average-ability peers.

4.2. Method

Participants

In the current study, 113 children, 54 boys and 59 girls, participated, ranging in age from 7;1 to 8;9 years (M=7.90). The average-ability children (n=68) attended mainstream elementary schools, and those who were identified as gifted were enrolled in special settings for gifted and talented children in the western part of the Netherlands. Gifted children (n=45) were over-sampled and preliminary identification of giftedness took place on the basis of their enrolment in gifted education and qualitative judgements of parents and teachers regarding their giftedness. Schools participated on a voluntary basis, and written permission to participate was obtained from the children's parents and schools prior to participation. Six children dropped out in the course of the study, as they did not participate in each test session.

Design

The study utilized a 2 x 2 pre-test-post-test control group design with randomized blocks with Ability category (gifted versus average ability) and Condition (dynamic testing versus unguided practice) as variables (see Table 1). Blocking was based on the scores on the Raven Standard Progressive Matrices test (Raven, 1981), a visual inductive reasoning test, administered before the pre-test. All the children who had been identified as gifted had obtained Raven scores of at least the 90th percentile. Then, Raven scores were used, per Ability category, in order to ensure differences in initial reasoning ability were as small as possible across the dynamic testing and unguided practice conditions, to block children into the unguided practice (control static) testing condition or the dynamic testing condition. Children in the dynamic testing subgroups received training between pre-test 2 and post-test, whereas children in the unguided practice subgroups received an unrelated dot-to-dot control task of equal length between pre-test 2 and post-test.

Table 1. Overview over the design

		Dynamic testing (n=53)[1]		Unguided practice (n=60)	
		Gifted (n=22)	Average-ability (n=31)	Gifted (n=23)	Average-ability (n=37)
Prior to dynamic/static testing	Raven	x	x	x	x
	BRIEF	x	x	x	x
	BCST-64	x	x	x	x
Dynamic/static test	Pre-test 1	x	x	x	x
	Pre-test 2	x	x	x	x
	Dynamic training	Dynamic training	Dynamic training	Dot-to-dots control task	Dots-to-dots control task
	Post-test	x	x	x	x

[1] This study employed the same participants as in the study described in Chapter 3

The design included pre-test sessions 1 and 2 in order to enable comparisons between static and dynamic progression. During the pre-test sessions and the post-test, all children were only provided with short, general instructions and were not given any feedback. Administration of the instruments, including the training session, took approximately 20-30 minutes per session.

Materials

Raven. All participants were administered the Raven Standard Progressive Matrices Test (Raven, 1981) as a measure of their intellectual ability and a blocking instrument. The Raven test is a non-verbal intelligence test that measures fluid intelligence by means of multiple choice figural analogies. The Raven test results were shown to have a high level of internal consistency in several studies as shown by split-half-coefficients of r=.91 (Raven, 1981).

Berg Card Sorting Test-64 (BCST-64). The Berg Card Sorting Test-64 (Piper, Li, Eiwaz, Kobel, Benice et al., 2011), the shortened version of the BCST, was used to measure cognitive flexibility. The BCST is an open-source computerized version of the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948). Evaluation of the BCST-64 has shown a very strong relationship with the full version of the BCST (Fox, Mueller, Gray, Raber, & Piper, 2013) and the WCST (Piper et al., 2011), and is therefore considered an appropriate alternative to the WCST. The number of perseverative errors made during the administration of the BCST-64 were used as a measure of the participants' cognitive flexibility. Higher perseverative errors correspond with lower cognitive flexibility.

BRIEF. The teacher questionnaire of the Dutch version of the Behavior Rating Inventory of Executive Functions (BRIEF; Smidts & Huizinga, 2009) was utilized to obtain an approximation of the teachers' evaluation of children's metacognition. Scores on the BRIEF Metacognition Index were used to obtain the teacher's evaluation of each child's metacognition. Higher scores of the BRIEF are associated with more deviations from the norm, or impairment of executive functions. The Metacognition Index was found to have a high level of internal consistency (Cronbach's α=.95, Smidts & Huizinga, 2009).

Dynamic version of geometric analogies.

Pre-tests and post-test. The dynamic test used in this study was comprised of geometric visuo-spatial analogies of varying difficulty of the type A:B::C:D (see Figure 1 for an example of a difficult analogy item), Both the pre-tests, and the post-test consisted of 20 items of various difficulty, part of a test battery originally created by Hosenfeld, Van den Boom, and Resing (1997), and adapted by Tunteler et al. (2008). Six basic geometrical shapes were used in the construction of the analogies: squares, triangles, hexagons, pentagons, circles, and ovals. Each analogy was constructed by means of five possible transformations: changing position, adding or subtracting an element, changing size, halving, and doubling. The test was administered as an open-ended paper-and-pencil test, and children had to draw their answers.

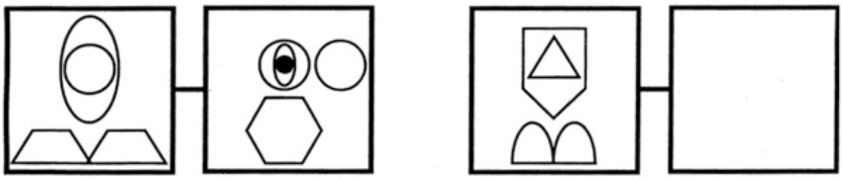


Figure 1. Example of a difficult analogy item.

The pre-tests and post-test, parallel sessions with different, but equivalent analogy items, were comprised of 20 trials with varying difficulty. The test sessions were equivalent in terms of the numbers of different elements, and transformations used for each analogy item, as well as the order in which the items were presented in relation to their difficulty level. The children received minimal instructions only in the two pre-tests and the post-test, as they were told that they had to solve puzzles with different shapes. The test leader then asked the child which shapes had to be drawn in the fourth box to solve the puzzle.

Training. The training session consisted of 10 geometric analogies that were not used in either the pre-tests or the post-test. The training session was based on graduated prompts techniques (Campione & Brown, 1987; Resing, 2000; Resing & Elliott, 2011). The prompts were administered following a standardized protocol, and were provided hierarchically, from two very general metacognitive prompts to two concrete cognitive prompts tailored to each specific item (see Appendix). Prompts were given if a child could not solve the analogy independently. After each prompt, children were asked to draw the solution of the analogy, and check their answer. If, after the fourth prompt, a child had not solved the analogy correctly, the test leader modelled the correct answer for the child. After the four prompts had been provided, and/or the test leader had shown the correct answer, the children were asked to explain why they thought their answer was correct. Then, the test leader provided a correct self-explanation. A schematic overview of the training procedure is included in the Appendix.

General procedure

The children were tested once a week over a period of five consecutive weeks. All tests and questionnaires part of the present study were administered following standard, protocolled instruction. At the beginning of the pre-tests, training session, and post-test, the children were provided with the six geometrical shapes used in the analogies, and in cooperation with the test leader named each shape, after which the test leader asked the child to draw the shapes below the printed shapes, staying as close to the original as possible.

Scoring

Analogy items were scored on the basis of children's drawings, in combination with their verbal explanations. Some of the children experienced difficulties drawing the geometrical shapes. This did not, however, cause any problems in scoring, as each child had copied the shapes used in the analogies on the cover sheet, so in the vast majority of cases the test leader knew which shapes the child had attempted the draw. In the few cases that it was, on first instance, unclear to the test leader which shape(s) the child had drawn, the child would be asked to point out on the cover sheet which shapes were intended.

For each item, the number of transformations that the child had applied correctly in solving the analogy was scored. Each analogy item was constructed by means of 1, 2, 3, 4, or 6 transformations that the child had to apply correctly in order to accurately solve the item, adding up to a total of 59 transformations per test session. The total number of transformations applied correctly in solving the analogies was taken as the outcome variable for each test session (Resing, Bakker, Pronk, & Elliott, 2016).

In order to estimate coding reliability, the pre-test 1 data were scored by both the first author and a student assisting in data collection. An inter-rater reliability analysis was performed using Cohen's κ to determine the level of agreement between the two raters. The inter-rater agreement for the pre-test 1 correct transformations was found to be very good, as determined by $\kappa=.83$, $p<.0001$.

Analyses

Multilevel modeling was used to analyze the current data. Multilevel modeling capitalizes on the hierarchical structure of the data, allowing us to study relations among variables at different levels and across levels. We can simultaneously answer level-1 questions about within-person change, and level-2 questions about how these changes vary across children (Singer & Willett, 2003). In the current study, level 1 represented the repeated measurements of the number of correct transformations within children, and level 2 represented the variability between children. We followed a predetermined model building structure as proposed by Singer and Willett (2003); starting with two simple, unconditional models and including our time-variant and time-invariant predictors in the successive models. The predictors were: condition, Ability category, cognitive flexibility and metacognition. Two time-invariant predictors, metacognition and cognitive flexibility, were mean centered to improve interpretation (Singer & Willett, 2003).

R (R Development Core Team, 2014) was used to fit the models. The fit of all models was compared using the likelihood ratio test (LRT) and two fit indices:

Akaike's Information Criterion (AIC) and the Schwarz's Bayesian Information Criterion (BIC). The likelihood ratio test follows a χ^2-distribution where the degrees of freedom are equal to the difference in the number of estimated parameters between the models. The LRT compares the "log likelihood" of two models and tests whether they differ significantly. The AIC and BIC are ad hoc criteria that are also based on the log likelihood statistic. The AIC and BIC statistics can be compared for all pairs of models, whether the models are nested within one another or not (Singer & Willett, 2003). These indices use a penalty function based on the number of parameters so that the more parsimonious model is favoured. A lower AIC and BIC value indicates a better fit of the model (Singer & Willett, 2003). All the discussed models were fitted using the Full Maximum Likelihood estimation. Most of the models differed in their fixed parts, and therefore deviance based on FML was needed to be able to compare the successive models (Singer & Willett, 2003).

4.3. Results

Before analysing the data for our research questions, one-way analyses of variance were conducted separately for each Ability category to evaluate possible differences between children in the two experimental conditions. The total Raven scores, pre-test 1 number of correct transformations, and age in months were used as dependent variables, and Condition with two levels (dynamic testing versus unguided practice) as independent variable. The findings for the gifted children revealed no significant differences in Raven scores (p=.53), pre-test 1 correct transformations (p = .40), nor in age (p=.52) between the dynamic testing and unguided practice conditions. Similarly, for the average-ability children no significant differences were found in Raven scores (p=.61), pre-test 1 correct transformations (p = .85), nor in age (p=.98) between the children in the two experimental conditions. We also examined possible differences between the gifted and average-ability children. The gifted children outperformed their peers on both the Raven scores, and the pre-test 1 correct transformations (for both measures, p<.001), but no significant differences were found in age (p=.31). Descriptive statistics of all measures used in the current study, per condition and Ability category are provided in Table 2.

Table 2. Mean scores and standard deviations of Raven scores, pre-test 1, pre-test 2, post-test correct transformations, cognitive flexibility and metacognition per condition and ability group

		Gifted		Average-ability	
		Dynamic testing	Unguided practice	Dynamic testing	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	39.14	41.96	29.16	28.43
	SD	15.13	9.26	13.56	15.77
Pre-test 2	M	46.86	53.74	43.52	41.03
	SD	17.62	4.05	13.40	18.27
Post-test	M	54.59	53.91	52.77	41.68
	SD	9.63	5.97	7.14	18.14
Cognitive flexibility	M	11.36	12.87	9.81	13.84
	SD	5.14	7.43	5.53	7.79
Metacognition	M	59.91	61.61	59.47	60.30
	SD	15.68	20.28	17.21	15.42

We conducted growth curve analyses (MLA) to model growth in the number of correct transformations. Table 3 presents the parameters and fit indices of the models. We first fitted the unconditional means model (intercept-only model) to acquire the random effects. The unconditional means model (Model 1) revealed a significant intercept effect (p<.001). We examined the intra-class correlation coefficient (ICC) as a measure of dependence; it describes the proportion of outcome variance that lies between persons in the population (i.e. the cluster structure of the data). As indicated by the intra-class correlation coefficient (ICC), of the total variation in the number of correct transformations, 54.38% could be attributable to differences between children. This finding revealed that the observations were not independent, and indicated that there was systematic variation in the outcome measure (transformations) worth exploring, both for the within-level and between-level variance, reinforcing the choice of multilevel modelling.

In Model 2 (the unconditional growth model), we included our time predictor into the level-1 sub-model in order to explain the remaining within-child variance (117.8). The estimated rate of change in the number of correct transformations for an average participant was 8.13 (p<.001); children generally improved in the number of correctly applied transformations. A negative covariance (-0.56) was found between the slope and intercept. This indicated that children using fewer correct transformations at pre-test 1 increased their number of correct transformations slightly faster across test sessions than children with a higher number of correct transformations at pre-test 1. Variance components revealed remaining variance in the number of correct transformations both between, and within, children. Extending the model by adding other predictors could possibly reduce this variation.

Model 3 included Condition as an explanatory variable for the number of correct transformations. Result of the likelihood ratio test (LRT) showed that model fit improved ($X^2(1)=5.46$, $p=.02$). Children of the unguided practice group had, on average, an estimated rate of change of 7.31. Therefore, these children generally increased their number of correct transformations across test sessions. A positive fixed effect for Condition (training versus unguided practice) of 3.51 revealed that the dynamic training session influenced the performance of the children. In accordance with our expectation, those who received a dynamic training session improved more in the number of correct transformations from pre-test 2 to post-test than the children in the unguided practice condition.

In Model 4 we included Ability category, gifted versus average-ability, as a predictor for initial status. Model 4 provided a better fit to the data compared to Model 3 ($X^2(1)=10.82$, $p=.001$). Children's Ability category was found to be related to the number of correct transformations at pre-test 1 as shown by a significant main effect of Ability category (8.23). Specifically, children with higher intellectual ability scored, on average, higher on pre-test 1 than average-ability peers. Model 5 showed that Ability category was also a significant predictor for children's rate of change, as indicated by a significant interaction of Ability category and Time. Model fit improved ($X^2(1)=4.96$, $p=.03$). The estimate (-2.21) revealed that average-ability children improved more in the number of correct transformations over time than gifted children.

In Model 6 we examined whether the dynamic training session had different benefits for gifted and average-ability children. We included the interaction effect of Ability category and Condition, which did not improve model fit ($X^2(1)=1.75$, $p=.19$). No significant difference was found in dynamic training benefits for gifted and average-ability children, as revealed by the non-significant interaction effect (-3.85), indicating that gifted children did not show more progression in the number of correct transformations after training than their average-ability peers.

Model 7 showed no significant main effect of Cognitive flexibility; model fit did not improve ($X^2(1)=0.53$, $p=.47$). The non-significant interaction effect of Cognitive flexibility x Time in Model 8 ($X^2(2)=0.59$, $p=.75$) indicated that we could not support our expectation that children with higher levels of cognitive flexibility would show more progression in the number of correct transformations than their age-mates with lower levels of cognitive flexibility. Children with higher levels of cognitive flexibility did also not benefit more from the dynamic training session than children with lower levels of cognitive flexibility as shown in Model 9 ($X^2(2)=2.84$, $p=.24$). Furthermore, results of Model 10 showed that the progression paths of gifted children that had higher levels of cognitive flexibility were not steeper than those of their average-ability peers ($X^2(2)=2.47$, $p=.29$). The time-invariant predictor Cognitive flexibility was not included in the remaining models.

Model 11 included the main effect of Metacognition. A non-significant effect was found, however, model fit did improve after inclusion of the predictor ($X^2(1)=22.80$, $p<.001$). Results of Model 12 showed that children with higher scores on the Metacognition Index showed equivalent progression in the number of correct transformations across test sessions than their peers with lower scores on the Metacognition Index ($X^2(1)=2.97$, $p=.08$). In Model 13, we included the interaction effect of Metacognition and Condition, which led to an improvement in model fit ($X^2(1)=4.40$, $p=.04$). The estimate (0.149) showed that children with higher scores on the Metacognition Index benefited more from training than peers with lower scores. We included the three-way interaction between Condition, Ability category and Metacognition in Model 14. Results showed that the progression paths of gifted children that had higher levels of metacognition were not steeper than those of their average-ability peers ($X^2(1)=0.20$ $p=.66$).

In conclusion, Model 13 was shown to be the model that best fitted the data based on the LRT, and the AIC and BIC statistics. The dynamic sessions led to an improvement in the number of correct transformations the children used. No differences in dynamic training benefits for gifted and average-ability children were found. The average-ability children in the unguided practice condition did, however, show more improvement across test sessions than the gifted children in the unguided practice session. Cognitive flexibility did not influence children's progression over time and the improvement in the number of transformations after receiving the dynamic training. The progression paths did also not differ for gifted children with higher levels of cognitive flexibility and their average-ability peers. Metacognition did not influence progression in the number of correct transformations. Children with higher scores on the Metacognition Index, indicating lower levels of metacognition, showed more improvement in the number of correct transformations after the dynamic training than their peers with lower levels of metacognition. Lastly, the progression paths did not differ between gifted children who had higher levels of metacognition and their average-ability peers.

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

Model	Estimate*(SE)*	Deviance	AIC	BIC
1. Intercept only	42.89(1.26)**	2750.6	2756.6	2768.1
2. Time	8.13(0.51)**	2557.8	2569.8	2592.7
3. Condition	3.51(1.40)*	2552.3	2566.3	2593.1
4. Ability category	8.23(2.39)**	2541.5	2557.5	2588.1
5. Ability category x Time	-2.21(0.98)*	2536.5	2554.5	2589.0
6. Ability category x Condition	-3.85(2.82)	2534.8	2554.8	2593.1
7. Cognitive flexibility	-0.13(0.17)	2536.0	2556.0	2594.3
8. Cognitive flexibility x Time	0.02(0.07)	2536.0	2558.0	2600.0
9. Cognitive flexibility x Condition	0.34(0.21)	2533.7	2555.7	2597.8
10. Cognitive flexibility x Condition x Ability category	0.49(0.35)	2534.1	2556.1	2598.2
11. Metacognition	-0.03(0.07)	2513.7	2533.7	2571.9
12. Metacognition x Time	0.05(0.03)	2510.8	2532.8	2574.8
13. Metacognition x Condition	0.15(0.07)*	2509.3	2531.3	2573.3
14. Metacognition x Condition x Ability category	-0.06(0.14)	2509.1	2533.1	2578.9

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

In order to examine our final research question regarding potential differences in the instructional needs of gifted and average-ability children, we conducted a one-way ANOVA with two within-subjects factors (metacognitive and cognitive prompts) and one between-subjects (Ability category) factor with the number of prompts in each category as dependent variables. No significant differences were found in the number of metacognitive, $F(1,51)=2.27$, $p=.14$, or cognitive prompts, $F(1,51)=.17$, $p=.69$ across ability categories (see Table 4).

Table 4. Mean scores and standard deviations of the number of metacognitive and cognitive prompts received during training per Ability category

	Metacognitive prompts		Cognitive prompts	
	M	SD	M	SD
Gifted	11.91	2.14	2.41	4.47
Average-ability	12.87	2.39	2.90	4.29

4.4. Discussion

The current study explored the potential differential benefits of dynamic versus static testing of gifted and average-ability children, and focused on two aspects of executive functioning, cognitive flexibility and metacognition. First of all, our results showed that children who had unguided practice experience only, and children who were dynamically tested showed progression in the number of correct analogical transformations. When children were tested dynamically, however, their progression paths were shown to be more advanced, which supports previous findings (Resing, 2000; Stevenson et al., 2013, 2014). In this sense, our findings build upon earlier studies in which it was posited that dynamic testing of children reveals a more complete picture of their cognitive potential than static testing only (e.g., Elliott, 2003; Sternberg & Grigorenko, 2002).

Moreover, our findings indicated, as expected, that gifted children start at a higher ability point, and keep this advantage during following sessions. When looking into potential differences between gifted and average-ability children in relation to the nature of progression, in contrast to our expectations, it was found that, in general, the average-ability children showed more progression than their gifted peers. We cannot, however, discount that the gifted children in the current study might have experienced a ceiling effect in testing, which could have influenced the research results. If these children had indeed experienced a ceiling effect, we would then have expected them to show a differential need for instructions, which could not be supported by our data. Moreover, neither the original authors of the items used in the current study (Hosenfeld et al., 1997), nor others who have used these items (e.g., Tunteler et al., 2008) for children of the same age report on a ceiling effect. It must be mentioned, nevertheless, that it is not known whether any high-ability children participated in these studies. Therefore, this explanation requires further research.

Looking more closely into training benefits, it was revealed that the gifted and average-achieving children showed similar rather than different progression lines after training, whereas previous studies into dynamic testing of gifted children found that these groups of children differed significantly in their performance and progression (e.g., Calero et al., 2011; Kanevsky, 2000; Kanevsky & Geake, 2004). Although we cannot completely discount a potential ceiling effect, as described above, in the light of the fact that all groups of children progressed after training, our findings, ultimately, seem to suggest that dynamic testing might be better suited to reveal children's cognitive potential of all groups of children (Elliott et al., 2010), including those with above-average cognitive abilities.

We also examined the role that cognitive flexibility and metacognition play in progression in accuracy of analogical reasoning, and training benefits.

It could not be established that cognitive flexibility plays a role in progression of analogical reasoning or training benefits. A number of reasons can be identified for the unexpected results regarding cognitive flexibility. First of all, research into executive functioning amongst children is challenging. One important reason is the type of instruments used to measure executive functioning. It has been noted that performance-based tasks, such as the BCST-64 used in the current study, rarely measure one executive function only (e.g., Miyake, Friedman, Emerson, Witzki, Howerter et al., 2000). By definition, executive functions regulate various cognitive processes, including for instance visuospatial processing. Performance-based tasks measure these other processes as well, making measuring just one executive function, in isolation, difficult (Viterbori et al., 2015). The developmental nature of executive functions in childhood should also be taken into consideration (e.g., Diamond, 2013; Kuhn, 2000). Moreover, it should be noted that the cognitive flexibility task used in the current study is a single measurement, static test, whereas learning potential measures are dynamic. Therefore, future studies could research this relationship further by utilizing a dynamic cognitive flexibility task, such as the dynamic Wisconsin Card Sorting Task (e.g., Boosman, Visser-Meily, Ownsworth, Winkens, & Van Heugten, 2014). These authors found that the dynamic executive functioning indices were significantly associated with cognitive functions, whereas the static indices were not.

It was, nonetheless, found that metacognition had an effect on the training benefits, but not on the progression from pre-test to post-test. Although it was expected that children with higher levels of metacognition would benefit more from training, we found the reverse. Children who, according to their teachers, had lower levels of metacognition benefitted more from training than their peers with higher levels of metacognition. This finding, once more, shows how dynamic testing can reduce test bias, and, in that way, lead to profound insights into how children learn (e.g., Elliott et al., 2010). Furthermore, the findings provide a first indication that a graduated prompts training procedure can, to a certain extent, compensate for lower levels of metacognition. This notion is particularly relevant considering Sternberg's (1998) assertion that metacognition is an important ability in the development of expertise.

Although it seems plausible that the graduated prompts technique used in the current study also helps improve metacognition, this tentative hypothesis cannot be asserted in the current study, and should be investigated using several measurements of metacognition. It must be noted that, although studies suggest that rating scales can be used successfully in order to obtain an approximation of children's executive functioning (Toplak, West, & Stanovich, 2013), using teacher ratings is a very indirect method of measuring metacognition. Future studies should therefore also focus on development and implementation of instruments that directly measure or predict executive functioning amongst young children.

Finally, we looked more closely into children's instructional needs during dynamic training. Contrary to what we expected based on previous literature (e.g., Calero et al., 2007; Kanevsky & Geake, 2004), we found no differences in the instructional needs of the gifted versus average-ability groups of children. Individual differences between children's need for instructions, both within and across ability categories, were, however, found, which is in line with previous studies (e.g. Resing, 2013; Sternberg & Grigorenko, 2002). Our study, moreover, suggests that children, regardless of whether they have high or average levels of cognitive abilities, can have a similar need for instructions in order to progress in learning. Of course, follow-up studies are required in order to investigate whether these findings are domain-specific or general.

In addition to the limitations mentioned above regarding measuring executive functioning and a potential ceiling effect, the current study encountered some other limitations. First of all, it is important to mention that we only used the Raven Standard Progressive Matrices as a measure of intellectual ability. Although the Raven test is known as a robust measure of intellectual ability (e.g., Jensen, 1998), we did not include other factors deemed important for cognitive and intellectual functioning, such as task commitment or creativity (e.g., Renzulli, 2005; Renzulli, & D'Souza, 2014). Moreover, we only investigated correct analogical transformations, while other factors have also been shown to be important in progression in analogical reasoning. Investigating strategy use, in particular, could lead to interesting findings considering the assumed relationship between strategy use and aspects of executive and intellectual functioning (e.g., Shore, 2000).

The results of the current study yield some important implications for educational professionals. In the context of the current study, it seems advisable to administer a dynamic rather than a static test when children's intellectual abilities are questioned, especially for children with lower levels of metacognition. Not only do our results underline the notion posited in a myriad of earlier work (e.g., Elliott et al., 2010; Resing, 2000, 2013; Sternberg & Grigorenko, 2002) that static testing does not always show a full picture of children's cognitive potential, our findings also indicate that children with different levels of intellectual ability, including those who have the potential to excel, can profit from dynamic testing, and, in particular, that children with lower levels of metacognition benefit more from training than their peers with higher levels of metacognition. Ultimately, the latter finding suggests that dynamic testing, in particular, may result in a more accurate view of the cognitive abilities of children with lower levels of metacognition.

Opponents of dynamic testing often argue that testing dynamically is more labour-intensive, and, thus, more expensive than testing statically. Nevertheless, as the children in the two ability categories showed progression after unguided repeated practice, and, more importantly, steeper progression lines after dynamic training, these findings suggest that gifted children also learn within the zone of proximal development (e.g., Calero et al., 2011). It seems that taking extra time to test these children more than once and administering a dynamic training session, helps them in unveiling their cognitive abilities, and, thus, is worth the extra investment.

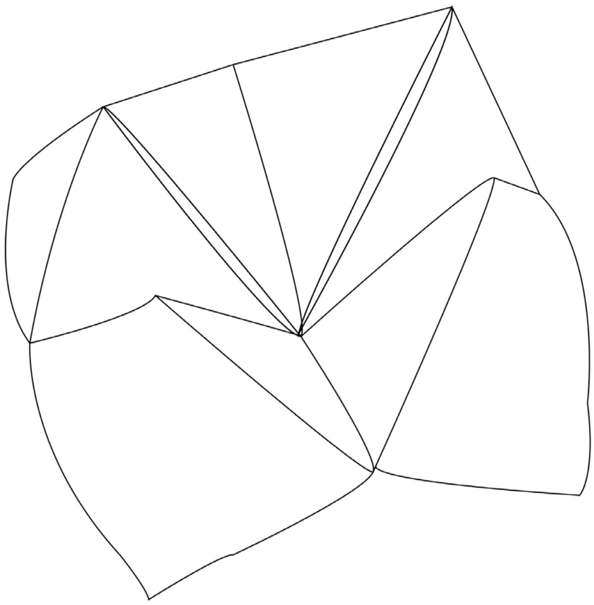
This notion becomes even more salient when taking into account that dynamic testing of children also provides insight into their instructional needs (e.g., Bosma & Resing, 2012). The results of the current study indicate that children of different levels of intellectual ability, including those with the potential to excel, can have a similar need for instructions, and can profit from similar help. Furthermore, our findings remind us that, when teaching high-ability children, these children do not, by definition, need less instruction or feedback than average-ability children, in order to show progression in learning. Just like any other children, some of these children can also profit from extra feedback or help so they can unveil their true cognitive potential. Finally, and most importantly, the results of the present study indicate that children, even those who have already achieved excellent results, can show learning progression when they are provided with the right instructions.

Appendix. Schematic overview of the graduated prompts training protocol

STEP	INSTRUCTION	INCORRECT ANSWER?	CORRECT ANSWER?
1	This is another puzzle with four boxes. Do you remember what we are going to do? *(have child provide an answer)* We are going to solve the puzzle by filling the empty box with the correct figures. Just draw the answer that you think is correct in the empty box *(have child draw the answer)*. Check whether you drew the correct answer *(have child check and correct answer if necessary)*	The picture you drew is great, but it is not entirely correct yet. I will help you, but try to find the correct answer with as little help from me as possible. We will start again after each try.	To step 5: Well done, that is the correct answer! Can you tell me why this this the correct answer? *[Test leader models correct self-explanation, as per the protocol, tailored to each item]*
2	How do we start? *(have child provide an answer)* First, have a good look at the figures in these three boxes *(point at A, B, C)* Do you now know the correct answer? Just draw the answer that you think is correct in the empty box *(have child draw the answer)* Check whether you drew the correct answer *(have child check and correct answer if necessary)*	Great picture! It is not entirely correct. I will help you some more.	
3	Have a good look at these boxes *[point at A and B]* What do you see? *[Have child provide an answer]* We see that A and B belong together. Do you know why? *[have child provide an answer]* *[Then explain the transformations from A → B according to protocol, tailored per item]* Do you now know the correct answer? Just draw the answer that you think is correct in the empty box *(have child draw the answer)* Check whether you drew the correct answer *(have child check and correct answer if necessary)*	You drew another beautiful picture. It is almost correct, so I will help you a little bit more.	
4	Now have a good look at this box *[point at C]* and this box *[point at A]* What do you see? *[Have child provide an answer]* We see that A and C look alike, but that they changed a little bit. Can you tell me why? *[Have child provide an answer]* *[Then explain the similarities between A and C, B according to protocol, tailored per item]* Do you now know the correct answer? Just draw the answer that you think is correct in the empty box (have child draw the answer) Check whether you drew the correct answer *(have child check and correct answer if necessary)*	What a beautiful picture. You can draw very well. It is not entirely correct; I will show you the correct answer *[test leader draws correct answer]* Can you tell me why this this the correct answer? *[Test leader models correct self-explanation, as per the protocol, tailored to each item]*	

CHAPTER 5

Changes over time and transfer of analogy-problem solving of gifted and non-gifted children in a dynamic testing setting



Bart Vogelaar
Wilma C. M. Resing

5

Vogelaar, B., & Resing, W. C. M. (under revision). Changes over time and transfer of analogy-problem solving of gifted and non-gifted children in a dynamic testing setting. *[Educational Psychology]*

Abstract
This study examined differences in transfer of analogical reasoning after analogy problem-solving between 40 gifted and 95 average-ability children (aged 9-10 years old), utilising dynamic testing principles. This approach was used in order to examine potential differences between gifted and average-ability children in relation to progression after training, and with regard to the question whether training children in analogy problem-solving elicits transfer of analogical reasoning skills to an analogy construction-task. Children were allocated to one of two experimental conditions: either children received unguided practice in analogy problem-solving, or they were provided with this in addition to training incorporating graduated prompting techniques. The results showed that gifted and average-ability children who were trained made more progress in analogy problem-solving than their peers who received unguided practice experiences only. Gifted and average-ability children were found to show similar progression in analogy problem-solving, and gifted children did not appear to have an advantage in the analogy-construction transfer task. The dynamic training seemed to bring about no additional improvement on the transfer task over that of unguided practice experiences only.

5.1. Introduction

Applying knowledge to a new context is an important necessity in order for gained knowledge and skills to be of use in everyday life outside the classroom context (Day & Goldstone, 2012), and is therefore one of the main aims of education. Groups of children have been found to differ in the extent to which they transfer learned knowledge and skills. Lower income students have, for example, been found to have more difficulty transferring knowledge and skills than their middle income peers (Alexander, Entwisle, & Olson, 2001). Other groups of learners have, in contrast, been found to have an advantage in transfer of learning, one such group being the gifted. Gifted children have long been thought to have an excellent ability to transfer learning to a new situation (e.g., Renzulli, Smith, White, Callagan, Hartman et al., 1997).

Considering the importance of transfer, it comes as no surprise that it has been studied for more than a hundred years (Engle, 2012). However, various studies have shown that eliciting transfer of learning to new contexts proves difficult (Bransford, Brown, & Cocking, 2001; Day & Goldstone, 2012; Gick & Holyoak, 1983), possibly due to its complex (Nokes-Malach & Mestre, 2013), and idiosyncratic nature (Kyllonen, Lohman, & Snow, 1984). Transfer has been noted for its potential to reveal important insights into children's potential for learning (Bosma & Resing, 2006; Ferrara, Brown & Campione, 1986). Therefore, in dynamic testing, transfer of newly acquired knowledge and skills is one of the measures used to gain insight into a child's potential for learning (Campione & Brown, 1987; Sternberg & Grigorenko, 2002; Tzuriel, 2007). In contrast with conventional static testing, dynamic testing is a form of testing that incorporates feedback and instruction, sometimes tailored to the individual, into the testing process (Elliott, 2003; Jeltova, Birney, Fredine, Jarvin, Sternberg et al., 2007), and is focused on the potential for learning, rather than on previously acquired skills and knowledge (Resing & Elliott, 2011).

In the present study, we applied dynamic testing principles to examine whether cognitively gifted and average-ability 9 and 10 year old children would show differential changes in analogy problem-solving, and differential patterns in their ability to transfer analogy problem-solving skills to an analogy-construction task. All the children in the present study received opportunities for unguided practice in analogy problem-solving. Half of the children, however, received an additional training in analogy problem-solving, which enabled us to investigate whether training would lead to more changes over time in problem-solving than unguided practice, and facilitate transfer of the learned skills.

Dynamic testing outcomes are assumed to provide a more detailed picture of a child's cognitive potential (Elliott, Grigorenko & Resing, 2010), strengths and weaknesses (Jeltova et al., 2007), than conventional, static testing procedures, such as intelligence or school aptitude tests (Elliott, 2003). This form of testing has been found to be especially beneficial for special populations, such as ethnic minority, or learning disabled children (Robinson-Zañartu & Carlson, 2013; Sternberg & Grigorenko, 2002). Perhaps due to these reasons, the large majority of research into dynamic testing has focused on special groups of children. Studies into dynamic testing of gifted children, however, are few. Findings of such studies have revealed that gifted children not only outperform their non-gifted peers, but also showed significantly more improvement (Calero, García-Martin, & Robles, 2011). Moreover, young gifted children were found to have a more extensive zone of proximal development, to learn new skills faster, and to be better at generalising new knowledge (Kanevsky, 1990, 2000).

Dynamic tests often employ inductive reasoning tasks (e.g., Ferrara et al., 1986; Resing, 2000; Resing & Elliott, 2011). Inductive reasoning is assumed to be related to a large variety of higher-order cognitive skills and processes (Csapó, 1997), including general intelligence (Klauer & Phye, 2008), problem solving (Richland & Burchinal, 2012), and applying knowledge and skills (Goswami, 2012). Analogical reasoning, a subtype of inductive reasoning, is considered to play a central role in cognitive development (Klauer & Phye, 2008; Pellegrino & Glaser, 1982), and develops significantly throughout childhood (e.g., Leech, Mareschal, & Cooper, 2008). Moreover, children's analogical reasoning ability can be characterised by large individual differences (e.g. Siegler & Svetina, 2002). Not only do older children perform better than younger children (Csapó, 1997), children with strong cognitive capacities, such as gifted and talented children, are also found to achieve higher scores on analogical reasoning tasks (e.g., Caropreso & White, 1994).

Several studies in the field of dynamic testing have revealed that training incorporating graduated prompting techniques, can lead to improvement in reasoning by analogy (Bosma & Resing, 2006; Ferrara et al., 1986; Stevenson, Heiser, & Resing, 2013). Graduated prompting techniques, as used in the present study, refer to a form of an intervention in which children are provided with prompts each time they make a mistake in problem solving. In the current study, prompts are tailored to each individual problem to be solved, and become more specific gradually, from metacognitive to cognitive prompts and modelling (Resing, 2000; Resing & Elliott, 2011). Graduated prompting techniques are used increasingly in combination with a pretest-training-posttest format, as in the present study, a specific form of dynamic testing that allows for structured measuring of children's progression in learning (e.g., Ferrara et al., 1986; Resing & Elliott, 2011; Resing, Bakker, Pronk, & Elliott, 2016).

Analogical reasoning involves defining and deciding that two problem-solving situations are similar, and, ultimately, successfully transferring previously problem-solving experiences to new situations that can be, partially, dissimilar. Unsurprisingly, reasoning by analogy is considered to be closely related to the ability to transfer (Gentner, Holyoak, & Kokinov, 2001); both require that one observes an analogy or similarity between two problems (Chi & VanLehn, 2012; Holyoak, 1984). In general, two factors have been proposed to play a role in transfer: the content, the exact problem that is being transferred (Barnett & Ceci, 2002, and the context (Klahr & Chen, 2011), which refers to the different domains from and to which the problem is being transferred. Researchers often distinguish in different types of transfer on the basis of the surface similarity of the base and target problem, including near versus far transfer (Mestre, 2005), and surface versus deep transfer (Forbus, Gentner, & Law, 1995). If the base and target share few surface similarities, and are thus less similar, the more cognitively demanding the process of transfer becomes (Gentner, Loewenstein, & Hung, 2007). Transferring effectively involves mastery of the task to be transferred (Siegler, 2006), and a deep, rather than surface understanding of the task at hand is required for deep transfer (Barnett & Ceci, 2002).

Several studies have shown that great variability exists in the extent to which children can transfer knowledge to new domains (e.g., Tunteler & Resing, 2010). It is often assumed that gifted children have a cognitive advantage, which enables them to transfer knowledge more efficiently than their non-gifted peers (e.g., Klavir & Gorodetsky, 2001; Zook & Maier, 1994). Research into the transfer ability of this group of children has revealed that on near transfer tasks gifted children's performance seems similar to their non-gifted peers (Carr, Alexander, & Schwanenflugel, 1996). In far transfer tasks however, gifted children were found to outperform their non-gifted peers (Geake, 2008; Kanevsky, 2000). Kanevsky (1990) reported that, after learning new strategies, gifted learners spontaneously transferred these strategies to new learning contexts. The underlying processes facilitating transfer in the gifted population are not yet fully understood, but Carr and colleagues (1996) suggest that gifted learners are more likely to transfer their acquired strategies to other domains, as they show an elaborate understanding, and make more use of complex strategies.

As eliciting transfer is challenging (e.g., Barnett & Ceci, 2002), a variety of studies have been carried out investigating whether training facilitating deep understanding or mastery of a task could promote transfer. Several studies have revealed that training children in solving inductive reasoning problems led to higher levels of generalising skills learned during training to similar and dissimilar problems in the same inductive reasoning domain (e.g., Harpaz-Itay, Kaniel, & Ben-Amram, 2006; Roth-van der Werf, Resing, & Slenders, 2002; Tzuriel, 2007). In the present study, we utilised a 'reversal' procedure to measure transfer. Having had practice opportunities, or practice opportunities in combination with a short training in analogy problem-solving, participants were asked to construct their own analogy items, similar to the ones they had solved before, which then had to be solved by the examiner (Bosma & Resing, 2006; Kohnstamm, 2014; Stevenson et al., 2013). As such, this task required a reversal of roles.

In order to promote transfer of problem-solving strategies practiced or trained, we kept the surface features of our analogy construction task similar to those of the open-ended visuo-spatial geometrical analogy items children solved before (Resing et al., 2016), assuming that children would use previously acquired knowledge and skills in their constructed analogies (Day & Goldstone, 2012). Previous research, however, has shown that despite these similarities in surface structure, the analogy construction task is a challenging and difficult task for children (Bosma & Resing, 2006; Tzuriel & George, 2009).

The present study had two main aims. Although consideration of the occurrence of transfer was our primary research aim, we were also interested in whether children's analogy problem-solving would improve differentially. Firstly, we sought to examine children's (differential) potential for learning. We expected that training by dynamic testing would lead to more change in children's analogy problem-solving than unguided practice only. We anticipated larger progression in accuracy scores of the children who were dynamically trained than the children wo received unguided practice only (Stevenson et al., 2013; Tunteler, Pronk, & Resing, 2008). We further anticipated that progression in accuracy would be larger for gifted than average-ability children, and there would be a significant interaction between session, condition and ability group for the accuracy scores (Calero et al., 2011).

We also considered the time it took children to complete all of the items of a test session. We expected that results for children in the unguided practice would show a decrease in completion time, but not for those in the dynamic testing condition, as we expected that training would lead children to spend time on strategic considerations (Resing, Tunteler, & Elliott, 2015). We also expected a significant three-way interaction of session x condition x ability group as to children's completion time, and hypothesised that gifted children would be more time efficient than their average-ability peers, considering they are assumed to be better in self-regulation (Calero, García-Martín, Jiménez, Kazén, & Araque, 2007).

Our second research question concerned the transfer of learned skills. As proposed by Clerc, Miller, and Cosnefroy (2014), in-depth assessment of transfer requires the measuring of both performance, as well as the degree of transfer achieved. Therefore, in the current study, we focused on both the transfer accuracy scores, as well as on the difficulty level of the analogies constructed by the children. We expected that, in comparison with children in the unguided practice condition, trained children would show higher levels of transfer accuracy scores; as well as difficulty levels of accurately constructed analogies (Resing, 1997; Roth-van der Werf et al., 2002). In addition, we expected that gifted children would show a higher degree of transfer than their average-ability peers (Geake, 2008; Kanevsky, 1990). We further explored whether children's transfer performance and degree of transfer could be predicted by their analogy problem-solving accuracy scores (Alexander & Murphy, 1999).

5.2. Method

Participants

In the present study, 135 children participated, 62 boys and 73 girls, ranging in age from 9 years and 3 months to 10 years and 11 months (M=10;10; SD=0;6). All the participants were born in the Netherlands, and attended either a mainstream primary school, or a special setting for gifted and talented children in the western part in the Netherlands. All 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 of parents and school was obtained for each child prior to participation in the current study. Nine children dropped out in the course of the study, as they did not participate in each test session.

Design

The study used a two-session (pre-test, post-test) repeated measures randomised blocking design with two treatment conditions: dynamic testing versus unguided practice (see Table 1). The children in the dynamic testing condition received two short training sessions between pre-test and post-test, whereas the children in the unguided practice condition did not receive any practice or training opportunities. Before the pre-test, the Raven Progressive Matrices Test (Raven, 1981) was administered to allocate children to the two treatment conditions. Only the children who had obtained a Raven percentile score of at least 90, were included in the "gifted" condition, the other children in the "average-ability" condition.

Raven scores were used to ensure that any differences in initial reasoning ability were as small as possible across the children in the dynamic testing and unguided practice conditions. Within the two ability groups, pairs of children with equal scores (blocking) were randomly assigned to the dynamic testing or unguided practice condition, resulting in four subgroups of children: gifted dynamic testing (n=22), gifted unguided practice (n=18), average-ability dynamic testing (n=47) and average-ability unguided practice (n=48).

Table 1. Overview of the design

Condition	Groups		Dynamic/static test				
		Raven	Pre-test	Training 1	Training 2	Post-test	Transfer
Dynamic testing	Gifted (n=22)	X	X	Dynamic training 1	Dynamic training 2	X	X
	Average-ability (n=47)						
Unguided practice	Gifted (n=18) Average-ability (n=48)	X	X	---	---	X	X

Materials

Raven Progressive Matrices Test. The Raven Progressive Matrices Test (Raven, 1981), a non-verbal intelligence test measuring fluid intelligence, was used as a blocking instrument. The Raven test results were shown to have a high level of internal consistency in several studies as shown by split-half-coefficients of r =.91 (Raven, 1981).

Dynamic test of analogical reasoning.

Pre-test and post-test. The dynamic test utilised visuo-spatial geometric analogies of the type A:B::C:?? of varying difficulty, part of a test battery developed by Hosenfeld, Van den Boom, & Resing (1997), and adapted for further use by Tunteler et al., (2008). Six basic geometrical shapes were used in the construction of the analogies: squares, triangles, hexagons, pentagons, circles, and ellipses (see Figure 1 for an example of a difficult analogy item). The original analogy test items were constructed by a maximum of five possible transformations: changing position, adding or subtracting an element, changing size, halving, and doubling.

As the original item-sets have been used for young children of various ages, but not for children from the age of nine, the items used in the current study were adapted by adding extra transformations, including rotation and colour. The test was administered as an open-ended paper-and-pencil test, and children were asked to draw their answers.

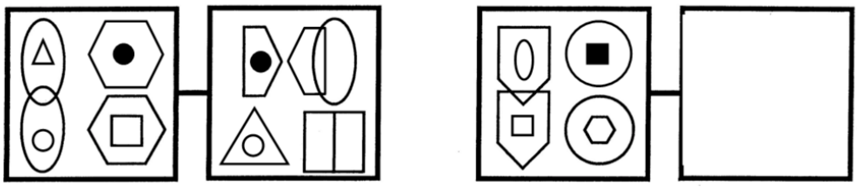


Figure 1. Example of a difficult analogy item

Both the pre-test and the post-test consisted of 21 items of varying difficulty. For pre- and post-test, parallel versions were constructed by keeping the difficulty levels of the items the same, as well as the order in which the items of varying difficulty were presented. Participants did not receive any feedback on or help with their given answers during the pre and post-test, but received minimal instructions that only specified the children had to solve puzzles by filling the empty square with the appropriate shapes. In our sample of participants, the pre-test was found to have high internal consistency (Cronbach's α=.85) for the accuracy scores.

5

Dynamic training. Two short training sessions each consisting of 6 new analogy items were administered between the pre-test and post-test to participants in the dynamic testing condition. The training sessions employed graduated prompting techniques used in earlier studies (e.g., Campione & Brown, 1987; Resing, 2000; Resing & Elliott, 2011). These involve the provision of a number of prompts when the child makes an error in problem-solving. All prompts were administered hierarchically: starting with four very general metacognitive prompts, followed by four specific cognitive prompts, tailor-made for each item. As each new prompt became progressively more specific, this procedure enabled measurement of the child's need for differing degrees of help in order to solve the problem presented. Both training sessions consisted of eight prompts in total, which were only administered after indication that a child could not solve the problem independently.

After each prompt, 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 seventh prompt had been administered, the examiner 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 shows a flowchart of the training procedure.

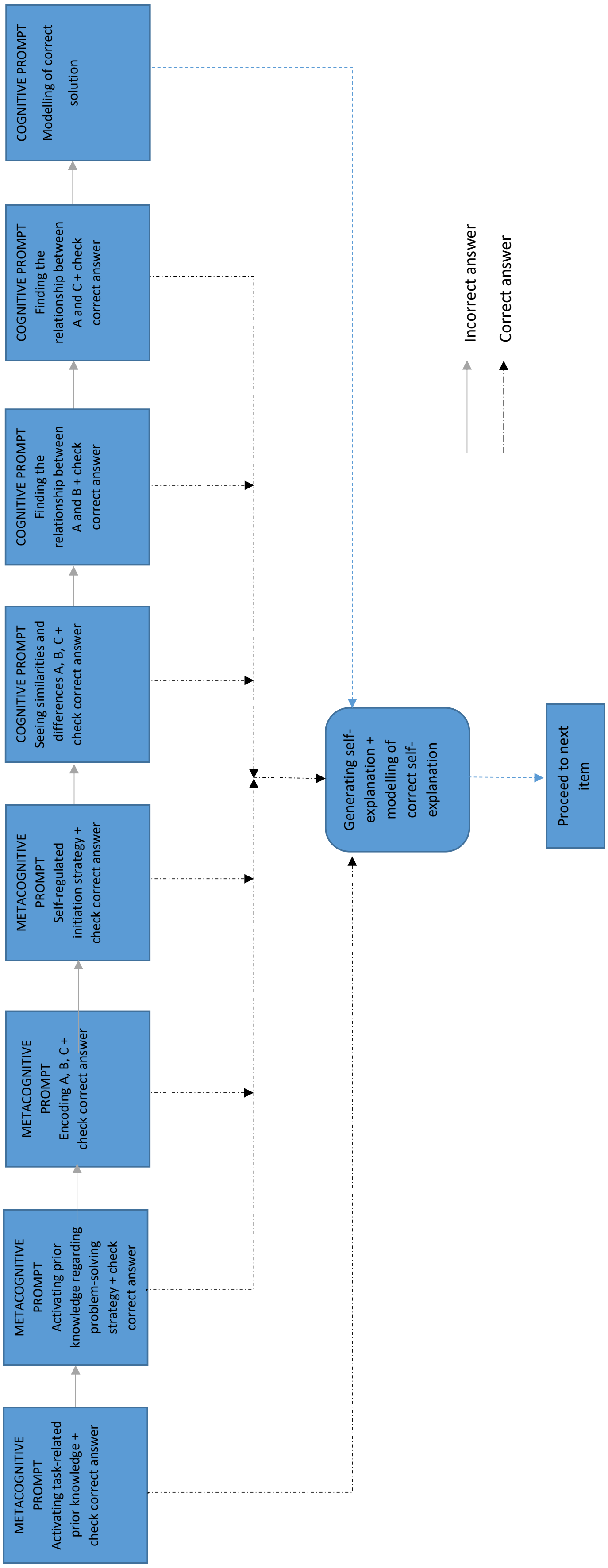


Figure 2. Flowchart of the graduated prompts training protocol.

Transfer: Analogy construction task. The ability to transfer previously practised or learned analogy problem-solving was measured by means of an analogy construction task. As part of this task, children were asked to construct their own analogy, so the examiner could solve it. In a sense, the roles were reversed, and the child became the teacher (Bosma & Resing, 2006). Participants were provided with four squares, similar to those used in the previous test sessions, but then empty, and instructed that they could utilise any of the geometric shapes they had seen in prior sessions, and to provide instructions to the examiner on how to solve the analogy. Deeper understanding of analogical reasoning principles is required to be able to construct a correct analogy (Harpaz-Itay et al., 2006). Children were asked to complete two reversal trials, and thus construct two analogies. For both tasks, the children were given short, general instructions only to enhance spontaneous problem-solving behaviour. After construction of the analogy item, the child had to ask the examiner to solve the item, and, on completion of the analogy by the examiner, then had to explain why this was the correct answer.

General procedure

Children were tested once weekly, in accordance with the schools' availability, over a period of six consecutive weeks. All parts of the present study were administered individually, following standard, protocolled instruction. At the beginning of the pre-test, training and post-test sessions, the children were provided with the six geometrical shapes used in the analogies, and, in cooperation with the examiner named each shape, after which the examiner asked the child to draw the shapes below the printed shapes, staying as close to the original as possible (Tunteler et al., 2008). It was assumed that this procedure helped activate children's prior knowledge, ensured that the test leader and child used the same terminology when addressing the geometric shapes, and, in doing so, facilitated the scoring procedure.

The solutions of the analogy items that the children had drawn during pretest, training and post-test were collected, and completion times of the pre-test and post-test were recorded. The analogies the children had constructed as part of the transfer task were collected, scored on
accuracy as well as on the number of transformations the item consisted of.

Scoring and analyses

The outcome variables of the pre-test and post-test sessions consisted of the accuracy score, the total number of correct items per session, and the completion time, the time (in seconds) it took each child to solve all the items of the pre-test and post-test.

We examined two outcome variables of the analogy construction task: transfer accuracy, and difficulty level. The first outcome variable was the sum of accurately constructed analogies (range 0-2). The second outcome measure was the transfer difficulty, calculated by means of the equation correctness of the analogy constructed (1/0) x the number of transformations used in the construction of the analogy (1-8; Mulholland, Pellegrino, & Glaser, 1980; Stevenson et al., 2013). The total difficulty level score (range 0-14) was calculated by adding the difficulty level scores of both items. Both outcome variables were considered to be ordinal, violating the assumptions of least-squares regression. Therefore, we conducted ordinal logistic regression analysis (Agresti, 2010). The regression analyses included the following predicting variables: condition, ability group, condition x ability group, post-test accuracy score, and condition x post-test accuracy score.

5.3. Results

Initial group comparisons

Prior to analysing our research questions, we evaluated possible differences between the two experimental conditions, and ability groups, respectively. The children in the two conditions did not differ in their age, or initial reasoning performance (Raven accuracy score). Children in the gifted and non-gifted groups also did not differ in age, but did in their initial reasoning performance ($p<.001$). We further evaluated possible differences in pre-test performance, and found no significant differences in accuracy scores, or in completion time between children in the two experimental conditions. Gifted and non-gifted children were found to differ on their accuracy scores ($p<.001$), but not on their completion time. Basic statistics for the measures used in the current study are provided in Table 2.

Table 2. Basic statistics of the analogical measurements, divided by condition and ability group

			Dynamic testing			Unguided practice		
			Gifted	Average-ability	Total	Gifted	Average-ability	Total
		N	22	47		18	48	
Raven	Accuracy	M	49.73	38.94	42.38	49.89	38.69	41.74
	scores	SD	2.51	5.90	7.15	2.78	6.18	7.41
Pre-test	Accuracy	M	10.77	5.91	7.46	9.83	7.08	7.83
	scores	SD	4.38	4.14	4.77	4.29	3.71	4.04
	Completion	M	1267.45	1207.52	1266.63	1331.09	1199.09	1235.09
	time	SD	302.80	564.37	494.54	367.80	471.58	446.87
Post-test	Accuracy	M	15.41	11.83	12.97	11.06	8.27	9.03
	scores	SD	3.42	4.89	4.76	5.13	4.15	4.57
	Solving-time	M	1055.05	1104.34	1088.62	1075.78	1027.42	1040.61
		SD	226.26	329.50	299.65	270.99	325.55	310.34
Transfer	Accuracy	M	1.77	1.66	1.62	1.94	1.50	1.70
	scores	SD	.43	.60	.65	.24	.72	.55
	Complexity	M	8.00	7.40	7.59	6.89	5.46	5.85
		SD	3.10	3.31	3.23	2.70	3.36	3.24

Analogy problem-solving

Our first research question concerned changes in analogy problem-solving. Changes over time were examined by means of two repeated measures ANOVAs with one within-subjects factor Session (Session 1-2), and two between-subjects factors Condition (dynamic testing versus
unguided practice) and Ability group (gifted versus average-ability). Children's accuracy scores and completion time at Sessions 1 and 2 were used as the dependent variables. In Table 3, the main and interaction effects of the repeated measures ANOVAs are provided. Results for accuracy scores revealed significant main effects of Session (p<.0001, η_p^2=.39), and, most important for answering our hypothesis, a significant Session x Condition (p<.0001, η_p^2=.20) interaction, but no significant Session x Ability group (p=.39), or Session x Condition x Ability group interaction (p=.36). Inspection of Table 3 led to the conclusion that, only partially in accordance with our hypotheses, dynamically tested children, irrespective of their ability group, showed significantly greater progression in solving analogies than control-group children. The slopes of the progression lines of the two dynamically tested groups of children did not significantly differ, indicating that children in both ability groups made comparable progress in accuracy although they started at different levels. The between-subjects effects of Ability group for

accuracy supported this finding, $F(1,131)=23.42$, $p<.0001$, $\eta_p^2=.15$.

Table 3. Results of the repeated measures ANOVAs for the accuracy scores, and completion time

	Wilks' λ	*F*	*p*	η_p^2
Accuracy scores				
Session	.62	82.09	<.0001	.39
Session x Condition	.80	32.39	<.0001	.20
Session x Ability group	.99	.76	.39	.01
Session x Condition x Ability group	.99	.84	.36	.01
Completion time				
Session	.86	21.75	<.0001	.14
Session x Condition	1.00	.49	.49	.00
Session x Ability group	.99	1.47	.23	.01
Session x Condition x Ability group	1.00	.03	.87	.00

A second aspect of children's analogy solving concerned the time they needed to complete all of the tasks of a test session. Although we expected that completion time would decrease for the children in the unguided practice condition, but not for the trained children, the repeated measures ANOVA showed only a significant main effect of session ($p<.0001$, $\eta_p^2=.14$), but no significant interaction effects (see Table 3). Contrary to our expectations, all groups of children showed a comparable decrease in their completion-time from pre-test to post-test.

These results led us to conclude that training leads to more improvement in accuracy than practice opportunities, as assumed, but, unexpectedly, that training and unguided practice both led to a decrease in solving time. Gifted and average-ability children seemed to differ in terms of the number of items solved correctly, as expected, with an advantage for those who were gifted, but, in contrast to our hypotheses, not in terms of change from pre-test to post-test, and training benefits in relation to both accuracy scores and completion time. Of course, individual differences in scores and changes in scores are large for children in both ability groups and conditions.

Transfer of analogy problem-solving

Our second research question concerned children's performance on the analogy construction transfer task. Ten children were unable to construct any accurate analogies, with eight children constructing items that were partial analogies, and two constructing items that were non-analogical. Out of 135 children, 27 and 98 children, accurately constructed either one, or two analogies, respectively (see Table 4). As a further exploration of the data, the children who had been able to construct accurate analogies were divided in three groups, based on the difficulty level of the analogies they constructed: low, medium, and high. Chi square tests revealed that children in the unguided practice condition constructed more analogies of a low difficulty level than the trained children ($\chi^2(1)=4.57$, $p=.03$), that trained children designed more difficult items than non-trained children ($\chi^2(1)=5.49$, $p=.02$), while the children who had constructed analogies of medium difficulty level were distributed evenly ($\chi^2(1)=.53$, $p=.47$). These findings revealed a first indication of the effect of training on transfer accuracy and difficulty level.

In addition, two Mann-Whitney U tests were conducted, one for each condition, to explore whether gifted and average-ability children were distributed evenly in relation to the difficulty level of the analogies they constructed. The results revealed that gifted and average-ability children were distributed evenly across the three difficulty level groups in the dynamic testing ($U=448$, $z=-.53$, $p=.60$), and the unguided practice condition ($U=392$, $z=.24$, $p=.81$).

Table 4. Children's transfer accuracy and difficulty level, divided by condition and ability group

	Dynamic testing			Unguided practice		
	Gifted	Average-ability	**Total**	Gifted	Average-ability	**Total**
Inaccurate analogies						
Non-analogical	0	0	**0**	0	2	**2**
Partial analogical	0	3	**3**	0	5	**5**
Accurate analogies						
Low difficulty (2-6 transformations)	4	7	**11**	8	18	**26**
Medium difficulty (7-9 transformations)	10	16	**26**	5	16	**21**
High difficulty (10-16 transformations)	8	20	**28**	5	8	**13**

An ordinal regression analysis was conducted to examine whether the number of accurately constructed analogies could be predicted by condition (dynamic testing versus unguided practice), ability group (gifted versus average-ability), and post-test accuracy. The results (see Table 5) revealed that the post-test accuracy score ($p=.001$) could significantly predict transfer performance. Although neither condition ($p=.18$), nor ability group ($p=.50$) significantly contributed to prediction, the condition x ability group interaction did (p=.04). The interaction was, contrary to our hypotheses, in the wrong direction.

A second ordinal regression analysis was conducted to examine whether transfer difficulty level could be predicted by condition, ability group and post-test accuracy. The findings indicated that only the post-test accuracy scores contributed significantly to prediction (p=.001). In sum, children's post-test accuracy score seemed to be a good predictor of transfer accuracy and difficulty level, regardless of training or ability group. Unexpectedly, however, the gifted children who received unguided practice seemed to outperform the gifted children who were trained in terms of transfer accuracy. An exploration of the quality of the constructed analogies suggested that differences between children were found mainly in the items of lower and higher difficulty group, with training seemingly facilitating construction of more difficult items.

Table 5. Results of the regression analyses for transfer accuracy and difficulty level (correct x transformations)

	b (SE)	Exp *(β)*	Exp *(β)*	χ^2	*P*
Accurate analogies [a]					
Condition	2.16 (1.59)	8.64	8.64	1.83	.18
Ability group	.48 (.71)	1.61	1.61	.45	.50
Condition x Ability group	-2.75 (1.33)	.06	.06	4.29	.04
Post-test accuracy score	.22 (.06)	1.25	1.25	11.82	.001
Condition x Post-test accuracy score	.09 (.10)	1.09	1.09	.73	.39
Transfer difficulty [a]					
Condition	.14 (1.07)	1.15	1.15	.02	.90
Ability group	.30 (.48)	1.35	1.35	.39	.53
Condition x Ability group	-.69 (70)	.50	.50	.98	.32
Post-test accuracy score	.16 (.05)	1.18	1.18	10.91	.001
Condition x Post-test accuracy score	-.01 (.07)	1.00	1.00	.01	.94

Note. [a] Ordinal logistic regression.

5.4. Discussion

The focus of the present study was two-fold. We examined gifted and average-ability children's progression in analogy problem-solving after dynamic training or unguided practice. We also focused on whether a dynamic training would facilitate children's performance and degree of transfer, and whether gifted and average-ability children show differences in their transfer accuracy and difficulty level.

We first looked into children's potential for learning. The children who, in addition to unguided practice experiences, also received dynamic training showed steeper progression in accuracy than the children who were not trained, indicating that testing children's ability dynamically shows a more complete picture of their cognitive potential than testing statically (see e.g., Robinson-Zañartu & Carlson, 2013). We also focused on potential differences between gifted and average-ability children in relation to changes in analogy problem-solving performance. Our findings suggest, as expected (e.g., Calero et al., 2011), that gifted children outperformed their average-ability peers in terms of accuracy in analogy problem-solving. They did not, however, show differential progression in accuracy or reduction in completion time, which is in contrast to earlier findings (e.g., Kanevsky, 1990, 2000).

In addition, children who were trained showed similar levels of reduction in time needed to solve the analogies to their peers who did not receive training. Although training seemed to lead to more advanced analogy problem-solving it did not lead to children spending more time on completing the tasks. In earlier research (Resing et al., 2015), it was posited that this might be due to children devoting more time to strategic considerations as a consequence of training. In the current study, however, children had to draw their own answers, which required substantial time and usage of motor skills. Completion time seemed to be dependent on other factors, such as children's fine motor skills. The reduction in completion time could, therefore, be ascribed to more familiarity with the task and an improvement in fine motor skills needed for the task, rather than other aspects, such as strategic considerations. In future studies, therefore, distinguishing between planning, and task execution time might lead to insights in relation to children's time allocation while solving analogy items, and whether gifted and average-ability children show differential patterns of time allocation and efficiency.

Our second main aim was to explore potential differences in gifted and average-ability children's transfer of practiced or learned skills. We utilised an analogy construction task (Bosma & Resing, 2006; Harpaz-Itay et al., 2006) in order to examine children's transfer accuracy and difficulty level. We expected that the transfer task would be difficult for the children, as it requires deep understanding of the task (e.g., Harpaz-Itay et al., 2006), and that at least some of the children would need training in order to facilitate this deep understanding. First of all, we found that the majority of children could construct an accurate analogy, in contrast with earlier studies in which more children were found to have difficulty with this task (e.g., Resing et al., 2016; Stevenson et al., 2013). The children participating in the current study were slightly older than in these previous studies, so we suggest that this is partly due to developments in their analogical reasoning (Csapó, 1997; Leech et al., 2008). Training, however, could not predict transfer accuracy or transfer difficulty level.

Further, children were divided in groups on the basis of the degree to which they could transfer, looking more closely at the difficulty level of the constructed analogy items. We found that the group of children that had constructed analogy items with a high difficulty level contained significantly more children who were dynamically trained, while the group that had constructed low difficulty analogy items consisted of significantly less dynamically trained children than children who had received unguided practice opportunities. These findings suggest that, for at least some children, training was necessary for them in order to construct the more difficult analogy items. Other children, however, did not need training to construct difficult items, reflecting individual differences between children in relation to their analogy problem-solving and construction skills.

One might assume that the group of children who constructed the more difficult items consisted mostly of the gifted children, but, contrary to our expectations, gifted children were not found to outperform their average-ability peers in transfer accuracy as well as difficulty level. In other dynamic testing studies, gifted children's performance was characterised by significantly more progression in performance (e.g., Calero et al., 2011) as well as higher transfer rates (e.g., Kanevsky, 1990), which led these authors to conclude that these children have a more extensive zone of proximal development. It must be noted that in the current study, children had not been formally identified as gifted by means of full scale IQ testing, but were identified on the basis of their parents' and teachers' judgements as well as their scores on the Raven test. Although the Raven test is considered a reliable measure of general intelligence, perhaps utilising a stricter cut-off score than the 90th percentile used in the current study, or taking into account other factors rather than just cognitive factors (see e.g., Kornilov, Tan, Elliott, Sternberg, & Grigorenko, 2012; Renzulli et al., 1997), would have led to more distinct differences in performance.

Moreover, it must be taken into consideration that the instructions of the transfer task did not specify that children had to construct complex analogies. Instead, instructions were kept to a minimum to elicit spontaneous problem-solving (e.g. Resing et al., 2016). Therefore, it cannot be disregarded that some of the gifted children were not motivated to construct difficult items, due to various reasons. Perhaps, some did not find the task sufficiently challenging to construct a highly complex item.

Clerc et al. (2014) provide an alternative explanation. These authors stated that self-regulation is strongly associated with the ability to transfer. They postulate that both children with low self-regulation as well as those with high self-regulation, such as gifted children, as for example demonstrated by Calero et al. (2007), can experience difficulty with transfer. According to them, good metacognition might hinder some children's transfer ability, as they do not want to use the strategy they have learned, before having fully mastered the cognitive processes necessary for utilising the strategy. A child's metacognitive knowledge relating to the transferral of a strategy or skill might be ahead of the child's actual ability to apply the strategy or skill. The fact that the gifted children who received unguided practice outperformed, in terms of transfer accuracy, their gifted peers who were trained lends some support to this explanation; perhaps training enhanced these children's metacognitive knowledge, while their actual ability to apply what they had learned in training was not yet at the same level, making these children unwilling to apply the strategies they have learned to a difficult item. Further research, with a larger sample of gifted children, ought to be conducted to further investigate these claims more thoroughly, taking into account, specifically, children's strategy use.

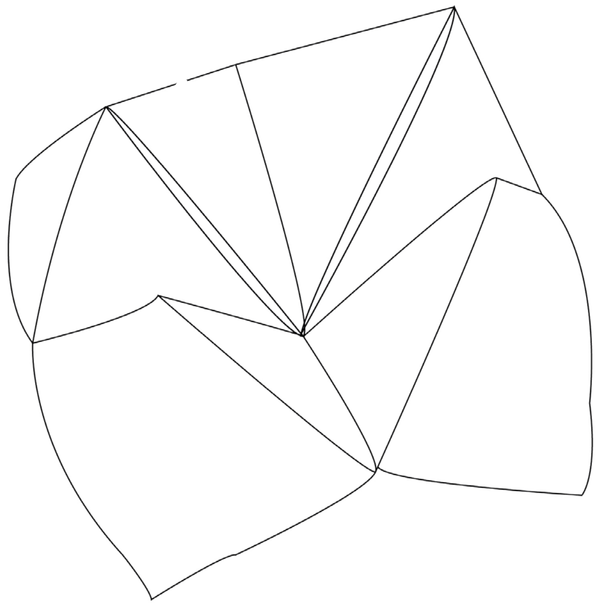
Clerc et al.'s (2014) explanation might also, in part, account for the fact that training, contrary to our expectations, could not predict transfer accuracy or difficulty level. Other explanations could be that the tasks were too difficult for some of the children to achieve deep understanding in a short time-frame (e.g., Tzuriel & George, 2009), or that the training employed in the current study was too short. In future studies, it might be useful to make the training more intensive, for example by increasing the number of sessions, or the number of items per training session (Resing et al., 2016; Tzuriel & George, 2009). More research, however, is necessary to investigate exactly what type of training is beneficial. Considering the individual differences portrayed by children in the current and in previous studies (e.g., Resing et al., 2015, 2016; Stevenson et al., 2013), children might benefit more from training that is more tailored.

Finally, the present study contributed to the existing research into transfer as we investigated both transfer accuracy, and complexity, and, thus, looked into both transfer performance and the degree of transfer obtained by the children (Clerc et al., 2014). These authors postulate that effectiveness of transfer can only be captured by measuring these two aspects. As children's accuracy in analogy problem-solving was found to predict both transfer performance, as well as transfer effectiveness, our findings support the notion that deeper understanding of analogy problem-solving is required for successful analogy construction (Harpaz-Itay et al., 2006).

The findings from the present study also support Siegler's (2006), and Day and Goldstone's (2012) suggestions that mastery of a skill is a requirement for transfer to occur, especially at the deep level, and that transfer at the deep level is challenging for young children (e.g., Clerc et al., 2014; Resing et al., 2016). Only a number of children could construct more difficult items, the majority of whom had received training in analogy-solving. Children showed considerable individual differences in their progression in accuracy, as well as their performance and effectiveness of transfer, findings that could only partially have been captured by traditional, static testing. In that sense, it seems plausible that dynamic testing might be a valuable instrument in capturing the underlying processes involved in progression in performance, as well as transfer in relation to learned skills.

CHAPTER 6

General Discussion



6

The field of research into dynamic testing is dominated by studies looking into the cognitive abilities of special populations, such as children from diverse ethnic backgrounds, with learning and intellectual disabilities while studies examining the cognitive abilities of gifted children are far more scarce. In this thesis, dynamic testing principles were applied to examine analogy problem-solving of gifted and average-ability children, in order to detect potential differences between these two groups of children.

Potential differences in analogy problem-solving were examined between gifted and average-ability children in relation to age (Chapter 2), instructional needs (Chapters 2 and 4), and transfer of learned skills by means of an analogy construction task (Chapter 5). In addition, within these two groups of children, two aspects of executive functioning, metacognition and cognitive flexibility (Chapter 4), were investigated as aspects that might facilitate the development of expertise, as posited by Sternberg's (1999, 2001; Sternberg, Jarvin, & Grigorenko, 2011) Developing Expertise Model. Test anxiety was examined with the objective to identify an aspect that potentially prevents the unfolding of the development of expertise (Chapter 3).

In this chapter, a summary of the most important findings of the studies in this thesis is provided. These findings will first be discussed in terms of their theoretical and practical considerations in relation to giftedness. Then, limitations of the studies in this thesis will be discussed. Finally, some implications for educational practice, and future research will be considered.

Summary of findings

The main aim of the study described in Chapter 2 was to identify potential differences between 5-8 year old gifted and average-ability children in relation to their potential for learning and need for instruction. It was found, in general, that dynamic testing and unguided practice opportunities both led to improvement, but that dynamic testing led to more advanced progression in accuracy. In addition, gifted children outperformed their average-ability peers in relation to accuracy in analogy problem-solving at each stage of the dynamic test, but, contrary to the expectations, showed equivalent progression paths, benefitted to similar degrees of training, and revealed similar degrees of instructional needs, both with regard to the amount, and the type of prompts they had received. Moreover, in line with the hypotheses, younger children showed less progression in accuracy, and needed more prompts than their older peers.

Chapter 3 focused on differences between 7-8 year old gifted and average-ability children with regard to their progression in analogy problem-solving after unguided practice opportunities or a dynamic training, and the potential differential impact of text anxiety on progression. The results of this study again revealed that dynamic testing led to more advanced progression in accuracy than unguided practice. Compared with their average-ability peers, gifted children demonstrated higher mean scores in relation to initial analogy problem-solving accuracy, and accuracy after unguided practice or training, but showed equivalent progression paths. Test anxiety was found to influence the children's rate of change across all test sessions, and their improvement in accuracy after dynamic training with children with higher levels of test anxiety benefitting more from training. Counter to the expectations, gifted and average-ability children did not differ significantly in the extent to which test anxiety was associated with their progression in solving analogies.

In Chapter 4, it was investigated whether two aspects of executive functioning, cognitive flexibility and metacognition, would be related to progression, after unguided practice or a dynamic training, of the number of correct transformations in analogy problem-solving of 7-8 year old gifted and average-ability children. Potential differences in instructional needs of gifted and average-ability children were also examined. The results revealed that dynamic testing led to more progression in the number of correct transformations than unguided practice opportunities. Gifted children demonstrated higher mean scores in relation to the initial number of correct transformations when solving analogies, and after unguided practice or training, but did not demonstrate steeper progression paths than their average-ability peers.

In contrast to the expectations, cognitive flexibility and metacognition did not influence children's progression over time, and the progression paths of gifted and average-ability children with higher levels of cognitive flexibility and metacognition were equivalent. Cognitive flexibility was also not found to be related to training benefits, but children with lower levels of metacognition, as estimated by their teachers, demonstrated more improvement in the number of correct transformations after the dynamic training than their peers with higher levels of metacognition. Finally, gifted and average-ability children required similar amounts of prompts during the dynamic training.

The aims of Chapter 5 were two-fold. The first aim concerned children's potential for learning, and it was investigated whether 9-10 year old gifted and average-ability children would show differential progression of accuracy in analogy problem-solving after unguided practice or a dynamic training. Secondly, it was examined whether gifted and average-ability children would demonstrate significant differences in transfer of solving analogies to an analogy construction task (focusing on both correctly constructed analogies and the difficulty level of the analogy items), whether training would facilitate transfer, and whether children's analogy problem-solving accuracy scores could predict accurately constructed analogies, and the difficulty level. The results indicated that dynamic testing led to more improvement in accuracy than unguided practice, and that unguided practice and dynamic testing led to an equivalent decrease in solving-time. Gifted and average-ability children differed in accuracy scores at each test session, with an advantage for those who were gifted, but not in completion time. Moreover, gifted and average-ability children showed equivalent progression paths in accuracy and solving-time after unguided practice opportunities or training.

With regard to transfer, it was found that training could not predict the number of accurately constructed analogies, nor the difficulty level of the analogies constructed by the children. No differences were found for gifted and average-ability children between the number of accurately constructed analogies, nor the difficulty level of the items constructed by them. However, when comparing the children who were trained with those who had received practice opportunities only, it became clear that there were more trained children who had constructed items of a high difficulty level, and less trained children who had constructed items of a low difficulty level than children who had practiced analogy problem-solving only.

Theoretical and practical considerations

Dynamic versus static testing

The studies presented here supported the assertion of several authors that dynamic testing unveils a more insightful view of children's ability to learn than static testing (e.g., Elliott, 2003; Elliott et al., 2010; Resing, 2013; Robinson-Zañartu & Carlson, 2013; Sternberg & Grigorenko, 2002). Children were found to improve more in accuracy of analogy problem-solving (Chapters 2, 3, and 5), as well as in the number of correct transformations they applied when solving analogies (Chapter 4), but not in the time it took them to solve the test items (Chapter 5). Moreover, dynamic testing revealed significant individual differences in children's (progression in) test scores (Chapters 2, 3, 4, 5), instructional needs (Chapters 2, and 4), and transfer success, as measured by the number of correctly constructed analogies and effectiveness, as measured by the difficulty level of the items constructed (Chapter 5). These findings, however, seem irrespective of ability category (Chapters 2, 3, 4, 5), and age (Chapter 2). It was repeatedly revealed that after training both gifted and average-ability children demonstrated (equivalent) progression in analogy problem-solving. The findings that the gifted children showed progression after a training further suggests that using dynamic testing to assess the cognitive abilities of high potential children is useful, and, more importantly, leads to a more insightful view of their capabilities than using a static test only.

All studies in this thesis had a (pre-test)-pre-test-training-post-test design, with graduated prompting techniques. As prompts were administered hierarchically, i.e. ranging from metacognitive to cognitive prompts to modelling, that became more specific whenever a new prompt was provided, these procedures allowed for measuring the different degrees of help individual children needed in learning a new task (Resing & Elliott, 2011). In this sense, this training procedure provided information on children's instructional needs (Resing, 2013).

Taking into account previous findings, suggesting gifted children to be more responsive to feedback (Kanevsky, 1994), and have an advantage in self-regulation (Calero, García-Martín, Jiménez, Kazén, & Araque, 2007; Zimmerman, 1989), these children were expected, in terms of the quantity of feedback, to need less prompts, and, in terms of the quality of feedback, less specific help. However, in Chapter 2 (5 to 8 year olds), and in Chapter 4 (7 and 8 year olds), it was consistently found that gifted children and their average-ability peers had equivalent needs for instructions, both with regard to the number of prompts and the type of prompts and qualitatively. Nevertheless, significant individual differences were found in both the amount as well as in the type of prompts children needed during training, regardless of whether they were identified as gifted or average-ability.

Developing Expertise Model

As demonstrated by the findings of the studies that are part of this thesis, it seems that abilities can be considered as entities that are not stable, or fixed, but dynamic, which can be developed further given the right circumstances (e.g., Sternberg & Grigorenko, 2002). In the light of Sternberg's (1999, 2001; Sternberg et al., 2011) Developing Expertise Model, gifted children's potential for learning could, in fact, also be considered as developing expertise. According to Sternberg (2001), individuals are gifted if they have an unusual ability to "advance from abilities that are ready to be developed to those that are developed" (p. 2). Sternberg further states that children are continuously engaged in a process of developing expertise when learning new knowledge or skills. According to him, gifted children are those who exhibit extraordinary potential in one or several of the skills involved in developing expertise.

The present thesis examined several aspects that, according to Sternberg (1999; 2001; Sternberg et al., 2011), play a pivotal role in learning, and, ultimately, the development of expertise; two aspects of executive functioning, metacognition and cognitive flexibility, considered part of the Developing Expertise Model, were examined. It was also studied whether another factor in his model, test anxiety, might prevent the unfolding of the development of expertise. Finally, transfer was also investigated within this framework.

Executive functioning. The provisional finding that children with lower levels of metacognition, as estimated by their teachers, progressed more after training, and thus, benefitted more from training than their peers with higher levels of metacognition tentatively suggests that children in this latter group have developed more expertise independently than those with lower levels, supporting to some degree Sternberg's model. It seems, further, that the dynamic training, to a certain extent, compensated children who had lower levels of metacognition, underlining once more the importance of testing children dynamically. Although a small effect, these findings seem to support Sternberg's (1999; 2001; Sternberg et al., 2011) assertion that metacognitive skills provide individuals with understanding and control of their cognition, which facilitates learning. Although it was expected that children with different levels of metacognition would show differential progression in analogy problem-solving after repeated practice and repeated practice in combination with a dynamic training, with an advantage for gifted children, this was not supported by the data. This finding was unexpected in the light of the fact that gifted children are often credited for having excellent metacognition (e.g., Shore, 2000), but could be explained by Sternberg's (2001, Sternberg et al., 2011) assertion that metacognitive skills are often domain-specific, and the teacher rating scale used to assess children's metacognition most probably provides a general estimation of children's metacognition, as demonstrated in the classroom. Moreover, in a recent study by Veenman, Bavelaar, De Wolf, and Van Haaren (2014) it was found that gifted learners are just as likely as their non-gifted peers to suffer from metacognitive deficiencies. According to these authors, gifted learners might rely primarily on their intelligence when performing tasks, as a result of which they do not feel the need to develop their metacognitive skills further.

A second aspect of executive functioning examined in Chapter 4 was cognitive flexibility, noted for its importance in the learning process (e.g., Diamond, 2013). Cognitive flexibility is considered by some researchers to be amongst the key components of cognitive adaptability, and is in that way critical to adaptive expertise, and problem solving (e.g., Haynie & Shephard, 2009; Moncarz, 2011). It is also assumed to be a component of creative thinking, one of the three sets of thinking skills identified by Sternberg (1999; 2001; Sternberg et al., 2011). Cognitive flexibility was measured in this study by means of a performance-based task. In this study, no support could be found for the hypotheses that cognitive flexibility would be related to the development of expertise, or training benefits. This could mean that cognitive flexibility does not play such an important role in analogy problem-solving, although it could also be related to the manner in which cognitive flexibility was measured. This is discussed in more depth under methodological considerations.

Test anxiety. Whereas the skills that are part of the Developing Expertise Model facilitate learning, other factors might, to some extent, hinder the learning of new knowledge and skills, and in that way, prevent a child from unfolding the further development of expertise. One of these aspects is test anxiety, examined in Chapter 3.

The findings of this study suggest that test anxiety can indeed have a negative impact on developing expertise, and that providing children with training in a certain skill might alleviate test anxiety levels (e.g., Bethge, Carlson, & Wiedl, 1982). In this respect, dynamic testing seems to have less bias towards children with test anxiety than static testing (cf. Meijer, 1996, 2001). Although several authors have proposed that gifted children may experience less (negative effects of) test anxiety (e.g., Zeidner & Shani-Zinovich, 2011), the results of Chapter 3 indicate that gifted and average-ability children experience similar levels of test anxiety, and that both groups of children show equivalent effects of test anxiety on their progression in analogy problem-solving.

6

Transfer. The ability to generalise learning to other contexts – known as transfer – was studied in Chapter 5. The findings of Chapter 5 lend support to the assumption that expertise in a skill improves the chances of successful transfer. In addition, children who achieved higher analogy accuracy scores at the post-test were found to demonstrate higher rates of accurately constructed analogy items (transfer success), as well as items of a higher difficulty level (transfer effectiveness), supporting Siegler's (2006) assertion that in order to transfer knowledge or skills successfully, mastery of the task at hand is required. In particular, this finding also supports Barnett & Ceci (2002)'s statement that deep transfer can only be achieved if an individual has reached deep rather than surface understanding of the task to be transferred (Barnett & Ceci, 2002). Support was also lent to the notion that transfer ability can be indicative of children's differential potential for learning, as significant individual differences were found between children, regardless of whether they were identified as gifted or average-ability (e.g., Bosma & Resing, 2006; Campione et al., 1985; Elliott et al., 2010).

Clerc, Miller, and Cosnefroy (2014) provided some rationale for the unexpected findings that training could not predict transfer accuracy or difficulty level. These authors postulated that self-regulation can interfere with transfer, making children with low or high self-regulation at risk for transfer difficulty. A child's metacognitive knowledge in relation to transferring a skill might be ahead of the child's actual ability to apply the skill. It was suggested in Chapter 5 that perhaps some of the gifted children who were trained were unwilling to apply the strategies they had learned in the training, as they might have felt their ability to apply what they had learned in training was not yet at the same level of their metacognitive knowledge in regards to analogy problem-solving.

Limitations

Solving analogy items

In this thesis, geometric analogy items were used to examine children's changes over time in analogy problem-solving. Accuracy scores, number of correct transformations, and solving time were used as indicators of children's analogy problem-solving skills. Potential bottom and ceiling effects were identified amongst the youngest and oldest study participants in Chapters 2, 3, and 4 (5-8 year olds). In order to avoid a ceiling effect amongst the 9 and 10 year old participants (Chapter 5), the difficulty level of the test sessions was increased by using items that contained more transformations only, and increasing the number of elements and transformations in certain items. Inspection of the mean scores as well as individual scores of children revealed larger mean differences between the two ability categories than in the studies with younger participants (Chapters 2, 3, 4), with none of the children reaching the test ceiling. The scores of the gifted and average-ability groups of children, however, demonstrated the same pattern as in the studies described in Chapters 2, 3, and 4 (see for example Figure 2 in Chapter 2), indicating that this pattern is fairly robust. A ceiling effect amongst the 7 and 8 year old children can, however, not be discounted as yet, and needs further examination. Future dynamic testing studies should be conducted amongst the same groups of children utilising more difficult tasks in order to confirm this.

Only children's quantitative analogy problem-solving performance was considered in the current thesis. In future studies, it would therefore be useful to investigate in more detail children's strategic considerations when solving these items. Studies amongst older participants suggested that novices and experts use different strategies when solving analogy problems (e.g., Ozkan & Dogan, 2013), and, utilising the framework provided by Sternberg's Developing Expertise Model, it would be interesting to investigate whether such differences are already apparent amongst primary school children, especially when considering individual differences demonstrated by children when solving analogies (e.g., Resing, 2013).

In order to do so, it might be worthwhile to computerise the tests used in this thesis. Earlier research has indicated that assessment mode, paper-and-pencil versus digital test version, did not influence children's strategy use when solving analogy items (Stevenson, Touw, & Resing, 2011). These authors found that administering the test of analogical reasoning digitally instead of on paper took significantly less time to administer and analyse, and allowed for registering additional test information. Computerising the analogy items could further enhance scoring uniformity in future studies.

The influence of executive functioning on progression in analogy problem-solving

The potential influence of executive functioning on analogy problem-solving was examined in Chapter 4. Since, as posited in the introduction, measuring the executive functioning of (young) children is considered challenging, both a performance-based task and a rating scale were used as executive functioning measures, in accordance with recommendations in the literature (e.g., Toplak, West, & Stanovich, 2013). As described above, the findings of this study could not fully support expectations based on previous findings. Potential reasons for this include the developmental nature of executive functions, and the nature of the tasks used in the study.

While it is known that executive functioning develops in childhood (Diamond, 2013), the exact nature of its development and underlying processes are not yet fully understood (Deák, 2004; Veenman, Van Hout-Wolters, & Afflerbach, 2006; Miyake & Friedman, 2012). What is clear, however, is that these issues make it complicated to measure executive functioning accurately, in particular in the light of the assumption that they are higher-order functions, which require assessment that involves complex paradigms and measures (Deák, 2004; Veenman et al., 2006; Miyake & Friedman, 2012).

It should further be noted that the performance-based task measuring cognitive flexibility, the BCST-64, is a single measurement, static test. Perhaps, utilising a dynamic task measuring cognitive flexibility, such as the dynamic Wisconsin Card Sorting Task (e.g., Boosman, Visser-Meily, Ownsworth, Winkens, & Van Heugten, 2014) in future studies would lead to different results. Likewise, metacognition was measured by means of a teacher rating scale. Research suggests that rating instruments do not always fully capture children's executive functioning (e.g., Sadeh, Burns, & Sullivan, 2012). Of course, individual differences between teachers when completing the rating form should also be taken into consideration. Likewise, as posited by Sternberg (2001; Sternberg et al., 2011), metacognitive skills are predominantly domain-specific, and a teacher rating scale provides an estimation of metacognition as demonstrated by the child in general in the classroom.

In sum, more research is needed to provide more information on the exact nature of executive functioning, its development, and underlying processes, as well as in relation to the instruments that can reliably capture different aspects and (sub)components of executive functions.

The influence of test anxiety on progression in analogy problem-solving

The influence of test anxiety on analogy problem-solving was examined in Chapter 3. Test anxiety was measured by means of the CTAS, a self-report questionnaire developed for children in grades 1-6 (Wren & Benson, 2004). Test anxiety was measured only once in this study, and as a result, previous findings that dynamic testing might reduce test anxiety (e.g., Bethge et al., 1982) could not be supported. While self-report measures are widely in the assessment of test anxiety (e.g., Wren & Benson, 2004), there are, however, some limitations associated with the use of self-report questionnaires, especially for young children, which ought to be mentioned. Social desirability (Galla, Plummer, White, Meketon, D'Mello et al., 2014), and memory distortions (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003) can affect the outcomes of a self-report questionnaire. Likewise, due to the fact that reading and language comprehension, memory, attention, abstract thinking and self-reflection are still developing significantly amongst young children, some authors question the use of self-report to assess the mental health of young children (e.g., Fallon & Schwab-Stone, 1994; Kuijpers, Otten, Vermulst, & Engels, 2014).

Therefore, in future studies investigating more closely the relationship between dynamic testing, and test anxiety scores, test anxiety could be measured before and after the dynamic test. It might be useful to combine self-report, with informant-report measures of test anxiety to obtain a more insightful, and objective view of children's test anxiety levels.

Characteristics of gifted children

In the current thesis, only cognitive aspects of the characteristics of gifted children were examined. Children were identified as gifted on the basis of parents' and teachers' nominations only (Chapter 2), or a combination of these with a percentile score of at least 90 of the Raven Standard Progressive Matrices Test, as a measure of their intellectual ability (Raven, 1981; Chapters 2, 4, and 5). The findings of this thesis suggest that, regardless of the identification process used, the gifted children showed similar patterns as the average-ability children, for example in relation to their progression in analogy problem-solving, and instructional needs. Although the Raven is considered a robust measure of general intelligence (e.g., Jensen, 1998), of course, there are several other factors that are assumed important in the cognitive and intellectual functioning of these children, such as task commitment or creativity (e.g., Renzulli, 2005; Renzulli & D'Souza, 2014).

Sternberg's Developing Expertise Model, for instance, also takes into account non-cognitive factors (e.g., Sternberg, 2001; Sternberg et al., 2011). The factors described in this model could be used in future studies when examining more closely both cognitive and non-cognitive factors that are associated with learning, and the development of expertise. The question as to how gifted children managed to achieve significantly higher performance scores than their average-ability peers has not been answered by the studies in this thesis. In future studies, it is therefore recommended to look more closely at the aspects of this model, and investigate to what extent these children demonstrate differences in the functioning of these elements, and in the direct and indirect relationships between these factors.

Practical implications

The notion that (static) cognitive test results do not always provide sufficient information for educational or pedagogical interventions has recently received more attention in research and practice (e.g., Compton, Fuchs, Fuchs, Lambert, & Hamlett, 2012; Fletcher, Stuebing, Barth, Denton, Cirino et al., 2011). This notion seems especially relevant in the light of recent changes in education policy leading to the implementation of inclusive education. Dynamic testing outcomes have been advocated as measures that reveal more insight into the processes and cognitive aspects that play a role in how children learn (e.g., Sternberg & Grigorenko, 2002; Resing, 2013), and in that sense, provide insightful information that can, for instance, serve as a starting point for the implementation of didactic interventions, and individual action plans (e.g., Jeltova, Birney, Fredine, Jarvin, Sternberg et al., 2007; Resing, 2013). The results found as part of this thesis resulted in a number of implications and recommendations for educational practice and future research, which are discussed in this section.

Assessment of children's cognitive abilities

The findings of the current thesis have consistently shown that testing children dynamically rather than statically results in a more accurate view of their cognitive potential. It was found that some children do not always show their full potential on a static test, and, thus, potentially underperform on a static test, which this thesis has suggested can be the result of deficits in metacognition, or test anxiety. Therefore, when children's cognitive abilities are tested for the purposes of decisions regarding the school level best suited for a child, it is recommended that dynamic tests are administered, especially when metacognitive deficits or test anxiety are suspected. Administering static tests might lead to underestimation of children's abilities, which, ultimately, could result in the loss of cognitive potential. Considering that the Dutch government aim to remain in the top five of knowledge-driven economies (Ministerie van Economische Zaken [Dutch Ministry of Economic Affairs], 2013), tapping into the potential of today's children is crucial. This recommendation seems especially valid for high-ability children, as today's high potential children are the scientists, politicians, directors, and entrepreneurs of the future.

Identification of gifted children

The previous recommendation also applies to the identification of giftedness. As discussed in Chapter 1, if a child is believed to be gifted, and in need of education that better suits his or her needs, it is common practice to test the cognitive abilities of these children statically. The outcomes of such testing procedures are then used to determine whether this particular child is eligible for participation in educational settings for the gifted and talented (e.g., Lohman & Gambrell, 2012). If, however, a strict IQ cut-off score of, for instance, 130, is used, a child that scores under 130, for example 129 or 128, might not be eligible for this type of education. This child is, however, potentially just as well suited for this type of education as a child scoring at or above the 130 IQ cut-off score. Therefore, it is recommended that, instead of focusing only on static test outcomes when considering whether a child is eligible for gifted education, it should also be considered how or why a child achieved a certain score, taking into account, for instance, the elements of the Developing Expertise Model. Moreover, such decisions should also be based on various information about a child's learning capabilities, instead of just one measure, including measures of potential for learning, and instructional needs. It seems valuable to make educational professionals more aware of the fact that gifted children do not always live up to their potential when they are being tested, especially regarding static tests, and that these children also learn within the zone of proximal development.

In conclusion, the findings discussed in this thesis question the idea that giftedness is a static entity, and that one simply "is" or "is not" gifted (Pfeiffer, 2011). It might be more worthwhile to think of giftedness as a more dimensional rather than a dichotomous concept (see e.g., Pfeiffer, 2011; Sternberg et al., 2011). It seems more valid to view giftedness not as a stable category that one simply "has" or "has not", but as an innate ability that is developmental and dimensional in nature, assuming there are different "levels" of giftedness (e.g., Sternberg, 2001; Sternberg et al., 2011) that, depending on several factors and circumstances may or may not be developed fully within an individual. These conclusions are in line with Subotnik et al. (2012), who acknowledge the developmental nature of giftedness, stating that in the beginning stages, giftedness might manifest as potential, and in later stages as achievement, and, fully developed, as eminence.

Tailoring to the educational needs of gifted children

The results of the current thesis underline that gifted children, just like non-gifted children, demonstrate significant individual differences when learning new skills, for instance in relation to their progression in learning, instructional needs, transfer ability, (influence on learning progression of) executive functioning, and levels of test anxiety. It is therefore crucial to ensure that gifted education incorporates possibilities for catering to individual learners' needs. Teachers and teacher educators should be made aware that gifted children cannot all be tarred with the same brush, and some of these children might even need extra attention or help to unfold their potential.

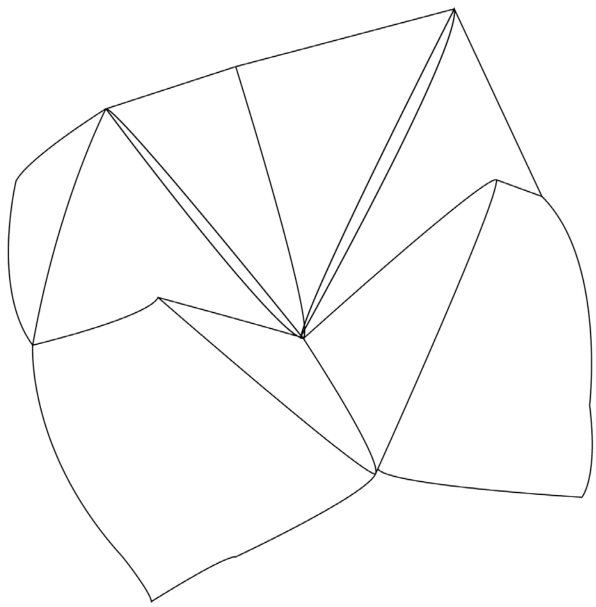
In practice, education for the gifted is often based on enrichment and/or acceleration principles (Gubbels, Segers, & Verhoeven, 2014; Hoogeveen, Van Hell, & Verhoeven, 2011; Schiever & Maker, 2003). Whereas these principles have proven to be effective for many gifted children, the results of this thesis suggest that a "one size fits all" approach does not benefit all gifted children. The findings of this thesis indicate that the instructional needs of gifted children are comparable to their average-ability peers in relation to the quantity and the type of instructions they need. This suggests, ultimately, that differentiation techniques in relation to instructional practice are necessary in gifted education, just like in other forms of education, for example by means of adaptive instruction (e.g., Heller, 1999).

Therefore, we advocate that education for the gifted, and other children alike, should be constructed on principles from the Developing Expertise Model (Sternberg, 1999, 2001; Sternberg et al., 2011), and the talent development framework. This latter framework emphasises "the deliberate cultivation of psychosocial skills supportive of high achievement, persistence, and creativity rather than leaving these to chance" (Olszewski-Kubilius & Thomson, 2015, p.54). According to these authors, the framework of talent development puts more emphasis on developing talent and potential, and, in that respect, provides more opportunities for tailoring to the needs of a more diverse range of children who are identified as gifted, including children with culturally and linguistically diverse backgrounds.

Future research

It was postulated in this thesis, and by other authors (e.g., Calero, García-Martín, & Robles, 2011; Kanevsky, 2000; Sternberg, 2001), that, just like other children, learning of gifted children occurs within the zone of proximal development. As revealed by the studies in this dissertation, gifted children's zone of proximal development is more advanced than that of their average-ability peers. These findings further seem to suggest that gifted children's learning, just like other children, can be characterised by the principles of the overlapping waves model posited by Siegler (1996). Three assumptions underpin this model: at any given time children have access to a variety of strategies that they can utilise to solve problems; they vary in which strategies they choose, suggesting that the strategies compete with each other; and the cognitive development of children is characterised by changes, occurring gradually, in relation to the frequency of utilisation of these strategies, as well as in the introduction of more advanced strategies, with the least effective strategies gradually disappearing. In line with this reasoning, it is recommended that future research focuses on the problem-solving processes of these children, investigating to what extent gifted children differ from average-ability children in their strategic choices. If conducted within the dynamic testing framework, such studies could shed more light on whether tapping into these children's zone of proximal development reveals differences in strategic functioning of these two groups of children.

REFERENCES



A

Agresti, A. (2010). *Analysis of ordinal categorical data.* Hoboken, NJ: John Wiley & Sons.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and *inequality:* A seasonal perspective. *Educational Evaluation and Policy Analysis, 23,* 171–191.

Alexander, P. A., & Murphy, P. K. (1999). Nurturing the seeds of transfer: A domain specific perspective. *International Journal of Educational Research, 31,* 561–576.

Alexander, P. A., White, C. S., Fuqua, J. D., Clark, G. D., Wilson, A. F., & Kulikowich, J. M. (1989). Development of analogical reasoning in 4- and 5-year-old children. *Cognitive Development, 4,* 65-88.

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology, 150,* 31–36.

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.

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition: Implications for the design of computer-based scaffolds. *Instructional Science, 33,* 367-379.

B

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1–29.

Balboni, G., Naglieri, J. A., & Cubelli, R. (2010). Concurrent and predictive validity of the Progressive Matrices and the Naglieri Nonverbal Ability Test. *Journal of Psychoeducational Assessment, 28,* 222–235.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128,* 612–637.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1-48.

Beckmann, N., Beckmann, J. F., & Elliott, J. G. (2009). Self-confidence and performance goal orientation interactively predict performance in a reasoning test with accuracy feed back. *Learning and Individual Differences, 19,* 277-282.

Bethge, H., Carlson, J. S., & Wiedl, K. H. (1982). The effects of dynamic testing procedures on Raven Matrices performance, visual search behaviour, test anxiety and test orientation. *Intelligence, 6,* 89-97.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and*

Psychopathology, 20, 899-911.

Boling, E., & Day, J. D. (1993). Dynamic assessment and giftedness: The promise of assessing training responsiveness. *Roeper Review, 16,* 110–113.

Boosman, H., Visser-Meily, J. M. A., Ownsworth, T., Winkens, I., & Van Heugten, C. M. (2014). Validity of the Dynamic Wisconsin Card Sorting Test for assessing learning potential in brain injury rehabilitation. *Journal of the International Neuropsychological Society, 20,* 1034–1044.

Bosma, T., & Resing, W. C. M. (2006). Dynamic assessment and a reversal task: A contribution to needs-based assessment. *Educational and Child Psychology, 23,* 81-98.

Bosma, T., & Resing, W. C. M. (2012). Need for instruction: Dynamic testing in special education. *European Journal of Special Needs Education, 27,* 1-19.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2001). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Budoff, M. (1987). Measures for assessing learning potential. In C. S. Lidz (Ed.). Dynamic assessment: *An interactional approach to evaluating learning potential* (pp. 173–195). New York, NY: Guilford Press.

C

Calero, M. D., García-Martín, M. B, Jiménez, M. I., Kazén, M., & Araque, A. (2007). Self-regulation advantage for high-IQ children: Findings from a research study. *Learning and Individual Differences, 17,* 328-343.

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.

Calero, M. D., Mata, S., Carles, R., Vives, C., López-Rubio, S., Fernández-Parra, A., & Navarro, E. (2013). Learning potential assessment and adaptation to the educational context: The usefulness of the ACFS for assessing immigrant preschool children. *Psychology in the Schools, 50,* 705-721.

Calkins, S. D., & Marcovitch, S. (2010). Emotion regulation and executive functioning in early development: Integrated mechanisms of control supporting adaptive functioning. In S. D. Calkins & M. A. Bell (Eds.), *Child development: At the intersection of emotion and cognition* (pp. 37–58). Washington, DC: APA Press.

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

Campione, J. C., Brown, A. L., & Ferrara, R. A. (1982). Mental retardation and intelligence. In R. J. Sternberg (Ed.) *Handbook of human intelligence* (pp. 392-490). Cambridge, United

Kingdom: Cambridge University Press.
Campione, J. C., Brown, A. L., Ferrara, R. A., Jones, R. S., & Steinberg, E. (1985). Breakdowns in flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence, 9,* 297–315.
Caropreso, E. J, & White, C. S. (1994). Analogical reasoning and giftedness: A comparison between identified gifted and nonidentified children. *Journal of Educational Research, 87,* 271-279.
Carr, M., Alexander, J. M., & Schwanenflugel, P. J. (1996). Where gifted children do and do not excel on metacognitive tasks. *Roeper Review, 16,* 198–204.
Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27*, 270-295.
Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.). *Advances in the psychology of human intelligence* (pp. 1–75). Hillsdale, NJ: Lawrence Erlbaum Associates.
Chi, M. T. H., & VanLehn, K. A. (2012). Seeing deep structure from the interactions of surface features. *Educational Psychologist, 47,* 177–188.
Clerc, J., Miller, P. H., & Cosnefroy, L. (2014). Young children’s transfer of strategies: Utilization deficiencies, executive function, and metacognition *Developmental Review, 34,* 378–393.
Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. (2012). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities, 45,* 79-95.
Csapó, B. (1997). Development of inductive reasoning: Cross-sectional measurements in an educational context. *International Journal of Behavioural Development, 20*, 609–626.

D

Dai, D. Y., & Chen, F. (2013). Three paradigms of gifted education: In search of conceptual clarity in research and practice. *Gifted Child Quarterly, 57,* 151-168.
Dai, D. Y., Swanson, J. A., & Cheng, H. (2011). State of research on giftedness and gifted education: A survey of empirical studies published during 1998 – 2010 (April). *Gifted Child Quarterly, 55,* 126-138.
Davey, G. C. L. (1994). Worrying, social problem-solving abilities, and problem-solving confidence. *Behaviour Research and Therapy, 32,* 327-330.
Davis, H. B., & Connell, J. P. (1985). The effect of aptitude and achievement status on the self-system. *Gifted Child Quarterly, 29,* 131-136.
Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist, 47,* 153–176.

Deák, G. O. (2004). The development of cognitive flexibility and language abilities. In R. V. Kail (Ed.), *Advances in Child Development and Behavior Volume 31* (pp. 271–327). Amsterdam/New York/Oxford: Elsevier.

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.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135–168.

E

Elliott, J. G. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review, 55,* 15-32.

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

Elliott, J. G., Lidz, C. S., & Shaughnessy, M. F. (2004). An interview with Joe Elliott and Carol Lidz. *North American Journal of Psychology, 6,* 349-360.

Engle, R. A. (2012). The resurgence of research into transfer: An introduction to the final articles of the transfer strand. *Journal of the Learning Sciences, 21,* 347–352.

Everson, H. T., Millsap, R. E., & Rodriguez, C. M. (1991). Isolating gender differences in test anxiety: A confirmatory factor analysis of the test anxiety inventory. *Educational and Psychology Measurement, 51,* 243-251.

F

Fabio, R. A. (2005). Dynamic assessment of intelligence is a better reply to adaptive behavior and cognitive plasticity. *Journal of General Psychology, 132*, 41-64.

Fallon, T., & Schwab-Stone, M. (1994). Determinants of reliability in psychiatric surveys of children aged 6-12. *Journal of Child Psychology and Psychiatry, 35,* 1391-1408.

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

Fletcher, J. M., Stuebing, K. K., Barth, A. E., Denton, C. A., Cirino, P. T., Francis, D. J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to reading intervention. *School Psychology Review, 40,* 3-22.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19,* 141–205.

Fox, C. J., Mueller S. T., Gray, H. M., Raber, J., & Piper, B. J. (2013). Evaluation of a short-form of the Berg Card Sorting Test. *PLoS ONE, 8(5)*: e63885.

Frasier, M. M., & Passow, A. H. (1994). *Toward a new paradigm for identifying talent potential.* Storrs, CT: University of Connecticut. The National Research Center on the Gifted and Talented, Research Monograph 94112.

G

Galla, B. M., Plummer, B. D., White, R. E., Meketon, D., D'Mello, S. K., & Duckworth, A. L. (2014). The Academic Diligence Task (ADT): assessing individual differences in effort on tedious but important schoolwork. *Contemporary Educational Psychology, 39,* 314–325.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*, 177-190.

Geake, J. G. (2008). High abilities at fluid analogizing: A cognitive neuroscience construct of giftedness, *Roeper Review, 30,* 187-195.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (Eds.). (2001). *The analogical mind.* Cambridge, MA: MIT Press.

Gentner, D., Loewenstein. J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development, 8,* 285-307.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1–38.

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

Granott, N. & Parziale, J. (2002). Microdevelopment: A process-oriented perspective for studying development and learning. In N. Granott & J. Parziale (Eds.), Microdevelopment: *Transition processes in development and learning* (pp. 1-28). Cambridge, United Kingdom: Cambridge University Press.

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology, 34,* 404-411.

Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities, 42,* 111–132.

Grigorenko, E. L., & Sternberg, R. J. (1998). *Dynamic testing. Psychological Bulletin, 124,* 75-111.

Gubbels, J., Segers, E., & Verhoeven, L. (2014). Cognitive, socioemotional, and attitudinal effects of a triarchic enrichment program for gifted children. *Journal for the Education of the Gifted, 37,* 378–397.

H

Hancock, D. R. (2001). Effects of test anxiety and evaluative threat on students' achievement and motivation. *Journal of Educational Research, 94,* 284-290.

Harpaz-Itay, Y., Kaniel, S., & Ben-Amram, E. (2006). Analogy construction versus analogy solution, and their influence on transfer. *Learning and Instruction, 16,* 583-591.

Haynie, M., & Shepherd, D. A. (2009). A measure of adaptive cognition for entrepreneurship research. *Entrepreneurship: Theory and Practice, 33,* 695-714.

Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational applications.* New York, NY: Cambridge University Press.

Heller, K. A. (1999). Individual (learning and motivational) needs versus instructional conditions of gifted education. *High Ability Studies, 10,* 9-21.

Holyoak, K. J. (1984). Analogical thinking and human intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol. 2* (pp. 199-230). Hillsdale, N.J.: Erlbaum.

Hoogeveen, L., Van Hell, J. G., & Verhoeven, L. (2011). Social-emotional characteristics of gifted accelerated and non-accelerated students in the Netherlands. *British Journal of Educational Psychology, 82,* 585-605.

Hopko, D., Crittendon, J., Grant, E., & Wilson, S. (2005). The impact of anxiety on performance IQ. *Anxiety, Stress & Coping, 18,* 17-35.

Hosenfeld, B., Van den Boom, D. C., & Resing, W. C. M. (1997). New instrument: Constructing geometric analogies for the longitudinal testing of elementary children. *Journal of Educational Measurement, 34,* 367-372.

Hosenfeld, B., Van der Maas, H. L. J., & Van den Boom, D. C. (1997). Indicators of discontinuous change in the development of analogical reasoning. *Journal of Experimental Child Psychology, 64,* 367-395.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications.* Second edition. New York, NY: Routledge.

Huizinga, M., Dolan, C. V., & Van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44,* 2017-2036.

I

Isquith, P. K., Crawford, J. S., Andrews Espy, K. A., & Gioia, G. A. (2005). Assessment of executive function in preschool-aged children. *Journal of Mental Retardation and Developmental Disabilities Research Reviews*, *11*, 209–215.

J

Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko E. L. (2007). Dynamic assessment as a process-oriented assessment in educational settings. *Advances in Speech-Language Pathology, 9,* 273-285.

Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko E. L. (2011). Making instruction and assessment responsive to diverse students' progress: Group-administered dynamic assessment in teaching mathematics. *Journal of Learning Disabilities, 44,* 381-395.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

K

Kanevsky, L. S. (1990). Pursuing qualitative differences in the flexible use of problem-solving strategy by young children. *Journal for the Education of the Gifted, 13,* 115-140.

Kanevsky, L. S. (1992). The learning game. In P. Klein & A. J. Tannenbaum (Eds.), *To be young and gifted* (pp. 204-241). Norwood, NJ: Ablex.

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

Kanevsky, L. S. (1995) Learning potentials of gifted students. *Roeper Review, 17*, 157-163.

Kanevsky, L. S. (2000). Dynamic assessment of gifted students. In K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed., pp. 283-296). Oxford, United Kingdom: Pergamon.

Kanevsky, L. S., & Geake, J. (2004). Inside the zone of proximal development: Validating a multifactor model of learning potential with gifted students and their peers. *Journal for the Education of the Gifted, 28,* 182□217.

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6,* 449–467.

Klahr, D., & Chen, Z. (2011). Finding one's place in transfer space. *Child Development Perspectives, 5,* 196–204.

Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning. A training approach. *Review of Educational Research, 78,* 85–123.

Klavir, R., & Gorodetsky, M. (2001). The processing of analogous problems in the verbal and visual-humorous (cartoons) modalities by gifted/average children. *Gifted Child Quarterly, 45,* 205-2015.

Kline, R. B. (2001). Brief cognitive assessment in children: Review of instruments and recommendations for best practice. In J. J. W. Andrews, D. H. Soklofske, & H. L. Janzen (Eds.). *Handbook of psychoeducational assessment: Ability, achievement, and* *behavior in children* (pp. 105-132). San Diego, CA: Academic Press.

Kohnstamm, G. A. (2014). *Jean Piaget: Children and the inclusion problem.* New Jersey, NJ: Transactions.

Kornilov, S., Tan, M., Elliott, J. G., Sternberg, R. J. & Grigorenko, E. L. (2012). Gifted identification with Aurora: Widening the spotlight. *Journal of Psychoeducational Assessment, 30,* 117-133.

Kornmann, J., Zettler, I., Kammerer, Y., Gerjets, P., & Trautwein, U. (2015). What characterizes children nominated as gifted by teachers? A closer consideration of working memory and intelligence. *High Ability Studies, 26,* 75–92.

Kreft, I., & De Leeuw, J. (2007). *Introducing Multilevel Modeling.* London, United Kingdom: Sage Publications.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9,* 178-181.

Kuijpers, R. C. W. M., Otten, R., Vermulst, A. A., & Engels, R. C. M. E. (2014). Reliability and construct validity of a child self-report instrument: The Dominic Interactive. *European Journal of Psychological Assessment, 30,* 40-47.

Kulesza, E. M. (2015). Searching nearest potential of children with intellectual disability – Dynamic assessment. *Practice and Theory in Systems of Education, 10,* 301-309.

Kyllonen, P. C., Lohman, D. F., & Snow, R. E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. *Journal of Educational Psychology, 76,* 130-145.

L

Leech, R., Mareschal, D., & Cooper, R. P. (2008). Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences, 31,* 357–414.

Lidz, C. S., & Elliott, J. G. (Eds.). (2000). *Dynamic assessment: Prevailing models and applications.* Amsterdam, The Netherlands: JAI/Elsevier Science.

Lidz, C.S. & Elliott, J. (2006). Use of dynamic assessment with gifted students. *Gifted Education International, 21,* 151-161.

Lidz, C. S., & Macrine, S. L. (2001). An alternative approach to the identification of gifted culturally and linguistically diverse learners. *School Psychology International, 22,* 74-96.

Lohman, D. F., & Gambrell, J. L. (2012). Using nonverbal tests to help identify academically talented children. *Journal of Psychoeducational Assessment, 30,* 25-44.

M

Matthews, D. J., & Foster, J. F. (2005). A dynamic scaffolding model of teacher development: The gifted education consultant as catalyst for change. *The Gifted Child Quarterly, 49,* 222-233.

McClain, M. C., & Pfeiffer, S. (2012). Identification of gifted students in the United States today: A look at state definitions, policies, and practices. *Journal of Applied School Psychology, 28,* 59–88.

McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *National Foundation for Educational Research, 21,* 89-101.

Meijer, J. (1996). *Learning potential and fear of failure.* Doctoral thesis, University of Amsterdam.

Meijer, J. (2001). Learning potential and anxious tendency: test anxiety as a bias factor in educational testing. *Anxiety, Stress and Coping, 14,* 337-362.

Mestre, J. P. (Ed.) (2005). *Transfer of learning from a modern multidisciplinary perspective.* Greenwich, CT: Information Age.

Ministerie van Economische Zaken [Dutch Ministry of Economic Affairs] (2013). *Uitwerking Regeerakkoord voor versterking kenniseconomie* [implementation of the governmental agreement regarding the strengthening of the knowledge-driven economy]. Retrieved from www.rijksoverheid.nl

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21,* 8–14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variables analysis. *Cognitive Psychology, 41*, 49–100.

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity, 9,* 35–45.

Moncarz, H. (2011). *The relationship between playing games and metacognitive awareness.* Doctoral dissertation, George Mason University, Fairfax, Virginia.

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.

Morris, L. W., & Liebert, R. M. (1969). Effects of anxiety on timed and untimed intelligence tests: Another look. *Journal of Consulting and Clinical Psychology, 33,* 240-244.

Moses, L. J., & Baird, J. A. (1999). Metacognition. In R. Wilson (Ed.), *Encyclopedia of cognitive neuroscience.* Cambridge, MA: MIT Press.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology, 12,* 252–284.

N

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

Nicpon, M. F., & Pfeiffer, S. I. (2011). High-ability students: New ways to conceptualize giftedness and provide psychological services in the schools. *Journal of Applied School Psychology, 27,* 293–305.

Nisbett, R. E. (2009). *Intelligence and how we get it: Why schools and cultures count.* New York, NY: Norton.

Nokes-Malach, T. J., & Mestre, J. P. (2013). Toward a model of transfer as sense-making. *Educational Psychologist, 48,* 184–207.

O

Olszewski-Kubilius, P., & Thomson, D. (2015). Talent development as a framework for gifted education. *Gifted Child Today, 38,* 49–59.

Ozkan, O., & Dogan, F. (2013). Cognitive strategies of analogical reasoning in design: Differences between expert and novice designers. *Design Studies, 34,* 161–192.

P

Passow, A. H., & Frasier, M. M. (1996). Toward improving identification of talent potential among minority and disadvantaged students. *Roeper Review, 18,* 198-202.

Pellegrino, J. W., & Glaser, R. (1982). Analyzing aptitudes for learning: Inductive reasoning. In R. Glaser (Ed.) *Advances in instructional psychology* (pp. 269–345). Hillsdale, NJ: Lawrence Erlbaum Associates.

Peña, E. D. (2000). Measurement of modifiability in children from culturally and linguistically diverse backgrounds. *Communication Disorders Quarterly, 21,* 87-97.

Persson, R. S. (2010). Experiences of intellectually gifted students in an egalitarian and inclusive educational system: A survey study. Journal for the *Education of the Gifted, 33,* 536–569.

Pfeiffer, S. I. (2011). Current perspectives on the identification and assessment of gifted students. *Journal of Psychoeducational Assessment, 30,* 3–9.

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

Piper, B. J., Li, V., Eiwaz, M. A., Kobel, Y. V., Benice, T. S., Chu, A. M., Olsen, R. H. J., Rice, D. Z., Gray, H. M., & Mueller, S. T. (2011). Executive function on the Psychology Experiment Building Language tests. *Behavior Research Methods, 44*, 110-123.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879–903.

R

R Development Core Team. (2014). *R: A Language and Environment for Statistical Computing (Version 3.12)* [Software]. Vienna, Austria: R Foundation for Statistical Computing Retrieved from http://www.R-project.org

Raven, J. (1981). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford, United Kingdom: Oxford Psychologists Press.

Reis, S. M., & McCoach, D. B. (2000). The underachievement of gifted students: What do we know and where do we go? *Gifted child quarterly, 44,* 152-170.

Reis, S. M., & Renzulli, J. S. (2010). Is there still a need for gifted education? An examination of current research. *Learning and Individual Differences, 20,* 308–317.

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60,* 180–184, 261.

Renzulli, J. S. (2002). Emerging conceptions of giftedness: Building a bridge to the new century. *Exceptionality, 10, 67–75.*

Renzulli, J. S. (2005). The three-ring definition of giftedness: A developmental model for promoting creative productivity. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 246–280). New York, NY: Cambridge University Press.

Renzulli, J. S., & D'Souza, S. (2014). Intelligences outside the normal curve: Co-cognitive factors that contribute to the creation of social capital and leadership skills in young people. In J. A. Plucker & C. M. Callahan (Eds.), *Critical issues and practices in gifted education: What the research says* (pp. 343–362). Waco, TX: Prufrock Press.

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., & Westberg, K. L. (1997). *Scales for rating the behavioral characteristics of superior students.* Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S., & Sytsma, R. E. (2008). Intelligences outside the normal curve: Co-cognitive traits that contribute to giftedness. In J. A. Plucker & C. M. Callahan (Eds.), *Critical issues and practices in gifted education: What the research says* (pp. 57–84). Waco, TX: Prufrock Press.

Resing, W. C. M. (1997). Learning potential assessment: The alternative for measuring intelligence? *Educational and Child Psychology, 14*, 68–82.

Resing, W. C. M. (2000). Assessing the learning potential for inductive reasoning in young children. In C. S. Lidz, & J. G. Elliott (Eds.). *Dynamic assessment: Prevailing models and applications* (pp. 229–262). Oxford, United Kingdom: Elsevier.

Resing, W. C. M. (2013). Dynamic testing and individualized instruction: Helpful in cognitive education? *Journal of Cognitive Education and Psychology, 12,* 81-95.

Resing, W. C. M., Bakker, M., Pronk, C. M. E., & Elliott, J. G. (2016). Dynamic testing and transfer: An examination of children's problem-solving strategies. *Learning and Individual Differences, 49,* 110–119.

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.

Resing, W. C. M., Touw, K. W. J., Veerbeek, J., & Elliott, J. G. (2016). Progress in the inductive strategy-use of children from different ethnic backgrounds: a study employing dynamic testing. *Educational Psychology, 36,* 1–19.

Resing, W. C. M., Tunteler, E., de Jong, F. M., & Bosma T. (2009). Dynamic testing in indigenous and ethnic minority children. *Learning and Individual Differences, 19,* 445-450.

Resing, W. C. M., Tunteler, E., & Elliott, J. G. (2015). The effect of dynamic testing with electronic prompts and scaffolds on children's inductive reasoning: A microgenetic study. *Journal of Cognitive Education and Psychology, 14,* 231-251.

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

Richland, L. E. & Burchinal, M. R. (2012). Early executive function predicts reasoning development. *Psychological Science, 24,* 87-9.

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.

Risemberg, R. & Zimmerman, B. J. (1992). Self-regulated learning in gifted students. *Roeper Review, 15,* 98-101.

Robinson, A., & Olly, J. L. (2014). *A century of contributions to gifted education: Illuminating lives.* New York, NY: Routledge.

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.

Roebers, C. M., Cimeli, P., Röthlisberger, M., & Neuenschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: an explorative study on their interrelations and their role for school achievement. *Metacognition and Learning, 2,* 151-173.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence, 53,* 118–137.

Roth-van der Werf, T. G. M., Resing, W. C. M., & Slenders, A. P. A. C. (2002). Task similarity and transfer of an inductive reasoning training. *Contemporary Educational Psychology, 27,* 296–325.

Ryan, T. G., & Coneybeare, S. (2013). The underachievement of gifted students: A synopsis. *Journal of the International Association of Special Education, 14,* 58-66.

S

Sadeh, S. S., Burns, M. K., & Sullivan, A. L. (2012). Examining an executive function rating scale as a predictor of achievement in children at risk for behavior problems. *School Psychology Quarterly, 27,* 236 –246.

Schiever, S. W., & Maker, C. J. (2003). *New directions in enrichment and acceleration.* In N. Colangelo & G. A. Davis (Eds.), Handbook of gifted education (pp. 163-173). Boston, MA: Allyn & Bacon.

Schneider, W. (2010). *Metacognition and memory development in childhood and adolescence.* In H. W. Waters, & W. Schneider (Eds.), Metacognition, strategy use and instruction (pp. 54–81). New York, NY: Guildford Press.

Scruggs, T. E., & Mastropieri, M. A. (1988). Acquisition and transfer of learning strategies by gifted and nongifted students. *Journal of Special Education, 22,* 153- 166.

Segool, K. N., Carlson, J. S., Goforth, N. A., Von der Embse, N., Barterian, J. A. (2013). Heightened test anxiety among young children: elementary school students' anxious responses to high-stakes testing. *Psychology in the Schools, 50,* 489-499.

Serpell, R. (2000). Intelligence and culture. In R. J. Sternberg (Ed.). *Handbook of intelligence* (pp. 549–580). New York, NY: Cambridge University Press.

Shore, B. M. (2000). Metacognition and flexibility: Qualitative differences in how gifted children think. In R. C. Friedman, & B. M. Shore (Eds.), *Talents Unfolding: Cognition and Development* (pp.167-187). Washington, DC: American Psychological Association.

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

Siegler, R. S. (2006). Microgenetic analysis of learning. In W. Damon, R. M. Lerner, D. Kuhn, & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language, Vol. 2* (6th ed., pp. 464□510). Hoboken, N.J: Wiley.

Siegler, R. S., & Crowley, K. (1991). *The microgenetic method: A direct means for studying cognitive development.* American Psychologist, 46, 606-620.

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

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York, NY: Oxford University Press.

Smidts, D. P., & Huizinga, M. (2009). BRIEF *Executieve Functies Gedragsvragenlijst:*

Handleiding [Manual for the BRIEF Executive Functions Behavioral Questionnaire]. Amsterdam, The Netherlands: Hogrefe.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London, United Kingdom: Sage Publications.

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

Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science, 26,* 127–140.

Sternberg, R. J. (1999). Intelligence as developing expertise. *Contemporary Educational Psychology,24,* 259–375.

Sternberg, R. J. (2001). Giftedness as developing expertise: A theory of the interface between high abilities and achieved excellence. *High Ability Studies, 12,* 159-179.

Sternberg R. J., Grigorenko E. L. (2001). *All testing is dynamic testing. Issues in Education, 7,* 137-170.

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

Sternberg, R. J., & Grigorenko, E. L. (2009). *Dynamic testing.* The nature and measurement of learning potential. Cambridge, United Kingdom: Cambridge University Press.

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

Stevenson. C. E., Bergwerff, C., Heiser, W. J., & Resing, W. C. M. (2014). Working memory and dynamic measures of analogical reasoning as predictors of children's math and reading achievement. *Infant and Child Development, 23,* 51-66.

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.

Stevenson, C. E., Hickendorff, M., Resing, W. C.M., Heiser, W. J., & de Boeck, P. A. L. (2013). Explanatory item response modeling of children's change on a dynamic test of analogical reasoning. *Intelligence, 41,* 157-168.

Stevenson C. E., Touw K. W. J., & Resing W. C. M. (2011). Computer or paper analogy puzzles: Does assessment mode influence young children's strategy progression?, *Educational and Child Psychology, 28,* 67-84.

Sub, A., & Prabha, C. (2003). Academic performance in relation to perfectionism, test procrastination and test anxiety of high school children. *Psychological Studies, 48,* 7-81.

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2012). A proposed direction forward for gifted education based on psychological science. *Gifted Child Quarterly, 56,* 176–188.

Swanson, H. L. (2006). Working memory and dynamic testing in children with learning disabilities. In S. J. Pickering (Ed.), *Working Memory and Education* (pp. 125–156). Burlington, MA:

Academic Press.
Swanson, H. L. (2010). Does the dynamic testing of working memory predict growth in nonword fluency and vocabulary in children with reading disabilities. *Journal of Cognitive Education and Psychology, 9,* 51-77.
Swanson, H. L. (2011). Dynamic testing, working memory and reading comprehension growth in children with reading disabilities. *Journal of Learning Disabilities, 44,* 358-371.
Swanson, H. L., & Howard, C. B. (2005). Children with reading disabilities: Does dynamic assessment help in the classification? *Learning Disability Quarterly, 28,* 17.
Swanson, J. D. (2016). Drawing upon lessons learned: Effective curriculum and instruction for culturally and linguistically diverse gifted learners. *Gifted Child Quarterly, 60,* 172–191.

T

Terman, L. M. (1925). *Genetic studies of genius Vol. 1: Mental and physical traits of a thousand gifted children.* Stanford, CA: Stanford University Press.
Threlfall, J., & Hargreaves, M. (2008). The problem-solving methods of mathematically gifted and older average-attaining students. *High Ability Studies, 19,* 83–98.
Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner Review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry, 54,* 131–143.
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.
Tunteler, E., & Resing W. C. M. (2007). Effects of prior assistance in using analogies on young children's unprompted analogical problem solving over time: A microgenetic study. *British Journal of Educational Psychology, 77,* 43-68.
Tzuriel, D. (2001). *Dynamic assessment of young children.* New York, NY: Kluwer Academic.
Tzuriel, D. (2007). Transfer effects of teaching conceptual versus perceptual analogies. *Journal of Cognitive Education and Psychology, 6,* 194–217.
Tzuriel, D., & George, T. (2009). Improvement of analogical reasoning and academic achievements by the analogical reasoning program (ARP). *Educational and Child Psychology, 29,* 71–93.
Tzuriel, D., & Kaufman, R. (1999). Mediated learning and cognitive modifiability. Dynamic assessment of young Ethiopian immigrant children to Israel. *Journal of Cross-Cultural Psychology, 30,* 359-380.

V

Van der Leeden, R. (1998). Multilevel analysis of repeated measures data. *Quality and Quantity, 32,* 15-29.

VanTassel-Baska, J. (2006). NAGC symposium: A report card on the state of research in the field of gifted education. *Gifted Child Quarterly, 50,* 339-341.

VanTassel-Baska, J., & Stambaugh, T. (2005). Challenges and possibilities for serving gifted learners in the regular classroom. *Theory into Practice, 44,* 211-217.

Veenman, M.V.J., Bavelaar, L., De Wolf, L., Van Haaren, M.P.G. (2014). The on-line assessment of metacognitive skills in a computerized environment. *Learning and Individual Differences, 29,* 123-130

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning, 1,* 3–14.

Vendetti, M. S., Wu, A. & Holyoak, K. J. (2014). Far-out thinking: Generating solutions to distant analogies promotes relational thinking. Psychological Science, 25928–933.

Viterbori, P., Usai, M. C., Traverso, L., De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: a longitudinal study. *Journal of Experimental Child Psychology, 140,* 38-55.

W

Wooding, G. S., & Bingham, R. D. (1988). Gifted children's response to a cognitive stressor. *Gifted Child Quarterly, 32,* 330-333.

Worrell, F. C., & Erwin, J. O. (2011). Best practices in identifying students for gifted and talented education programs. *Journal of Applied School Psychology, 27,* 319–340.

Wren, D. G., & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress, and Coping, 17,* 227 – 240.

Z

Zeidner, M. (1998). *Test anxiety: The state of the art.* New York, NY: Plenum Press.

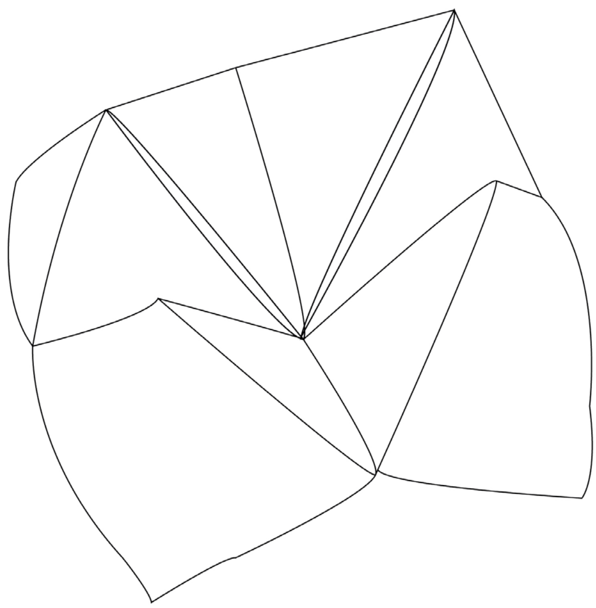
Zeidner, M., & Schleyer, E. J. (1999). Test anxiety in intellectually gifted school students. *Anxiety, Stress & Coping: An International Journal, 2,* 163-189.

Zeidner, M., & Shani-Zinovich, I. (2011). Do academically gifted and non-gifted students differ on the Big-Five and adaptive status? Some recent data and conclusions. *Personality and Individual Differences, 51,* 566-570.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81,* 329-339.

Zook, K. Z., & Maier, J. M. (1994). Systematic analysis of variables that contribute to the formation of analogical misconceptions. *Journal of Educational Psychology, 86,* 589- 600.

SUMMARY IN DUTCH



Samenvatting in het Nederlands

Als in het onderwijs verondersteld wordt dat een kind hoogbegaafd is, wordt veelal gebruik gemaakt van een (verkorte) conventionele intelligentietest om zijn of haar cognitieve capaciteiten te onderzoeken (Nisbett, 2009; Pierson, Kilmer, Rothlisberg, & McIntosh, 2012). Als onderdeel van zulke testen maakt het kind, na een gestandaardiseerde korte instructie, zelfstandig de testopgaven. Hoewel deze vorm van testen veel voordelen heeft, geven sommige onderzoekers aan dat het vaststellen van begaafdheid middels deze conventionele, statische, intelligentietesten niet voldoende is (Lohman & Gambrell, 2012; Worrell & Erwin, 2011). Niet alleen wordt aangenomen dat conventionele tests slechts gedeeltelijk inzicht bieden in het cognitief functioneren van kinderen (Elliott, 2003; Lohman & Gambrell, 2012), maar ook dat zij voor een groot gedeelte meten wat een kind in het verleden geleerd heeft. Voorgaande leerervaringen zijn echter niet altijd een goede voorspeller van het cognitief potentieel van een individu (Sternberg & Grigorenko, 2002).

Zo suggereert onderzoek bijvoorbeeld dat statische testresultaten nadelig kunnen uitvallen voor bepaalde kinderen, zoals diegenen die niet de mogelijkheid hebben gehad om optimaal kennis en vaardigheden op te doen in het onderwijs, of thuis. Te denken valt hierbij aan kinderen met een lage socio-economische status, een andere culturele achtergrond, bijzondere onderwijsbehoeften (Robinson-Zañartu & Carlson, 2013), of kinderen met testangst (Meijer, 1996, 2001). Bovendien ligt de nadruk bij conventionele, statische intelligentietests op de testresultaten en kunnen de psychologische processen die een rol spelen bij het leren niet of slechts indirect worden gemeten (Jeltova, Birney, Fredine, Jarvin, Sternberg et al., 2007). Onderzoekers en onderwijsprofessionals stellen dan ook vragen bij de praktische inzetbaarheid van deze vorm van testen bij didactische vraagstukken (Sternberg & Grigorenko, 2002).

In deze dissertatie staat een alternatieve manier van het testen van cognitieve capaciteiten centraal, het zogeheten dynamisch testen. In een dynamische test zijn feedback en/of hulp geïntegreerd in de testafname, waarbij deze vorm van testen een beeld verschaft van het leervermogen, of leerpotentieel, van een individu. Deze tests zijn gebaseerd op het gedachtegoed van, onder andere, Vygotsky en zijn concept van de zone van naaste ontwikkeling (Sternberg & Grigorenko, 2002). Deze zone kan worden beschouwd als het verschil tussen dat wat een individu zelfstandig, zonder hulp van anderen, kan bereiken (ook wel bekend als het actuele ontwikkelingsniveau) en dat wat een individu kan bereiken met hulp van een andere, capabelere persoon (ook wel bekend als het potentiële ontwikkelingsniveau).

Van oudsher zijn deze tests veelal gebruikt voor kinderen van speciale, voornoemde, populaties, zoals kinderen met een andere etnische achtergrond, of kinderen met leerproblemen. Het gebruik van dynamische tests voor kinderen met hoge intellectuele capaciteiten, zoals hoogbegaafde kinderen, is echter onderbelicht. Het eerste doel van deze dissertatie was dan ook meer inzicht te verkrijgen in de waarde van het gebruik van dynamisch testen voor het meten van de cognitieve capaciteiten van hoogbegaafde kinderen.

Een belangrijk uitgangspunt bij het dynamisch testen is dat cognitieve capaciteiten niet stabiel zijn, maar kunnen ontwikkelen (Sternberg, 1999, 2001; Sternberg & Grigorenko, 2002). Volgens deze denkwijze kan leren gelijkgesteld worden aan het ontwikkelen van expertise. De tweede doelstelling van deze dissertatie was dan ook het onderzoeken van enkele factoren die mogelijk een bevorderende, te weten metacognitie en cognitieve flexibiliteit, dan wel remmende, te weten testangst, rol kunnen spelen bij het verder ontvouwen van expertise.

In **Hoofdstuk 1** werden de theoretische en methodologische achtergronden besproken die ten grondslag liggen aan de studies die onderdeel zijn van deze dissertatie. De dynamische tests die gebruikt zijn in de studies onderdeel van deze dissertatie hadden alle een (pre-test)-pre-test-training-post-test ontwerp. Dit houdt in dat de deelnemende kinderen allereerst één of twee keer getest werden, de voormeting, zonder dat zij daarbij enige hulp of feedback ontvingen. Hierna werd de helft van de kinderen een training gegeven, gebaseerd op *graduated prompting* technieken, en de andere helft van de kinderen ontving een alternatieve controle-taak, waarbij zij alleen herhaald oefenden met de taak. Hierna werd bij alle kinderen de post-test, de nameting, afgenomen. De post-test bestond in alle studies uit een test die qua het type opgaven en de moeilijkheidsgraad ervan parallel was aan de pre-test(en). Een dergelijk onderzoeksontwerp maakt het mogelijk om op gestructureerde wijze van een individueel kind de vooruitgang te meten, inzicht te verschaffen in de individuele behoefte aan instructie, en zodoende een indicatie te verschaffen van het leerpotentieel. In een dergelijke test worden de score op de post-test, de vooruitgang van kinderen van pre- naar post-test alsmede het aantal en de type hints die kinderen hebben gekregen gezien als maten voor het leerpotentieel.

De training die de kinderen ontvingen is gebaseerd op het bieden van prompts, of hints, op het moment dat duidelijk wordt dat een kind een testopgave niet zelfstandig op kan lossen (Campione & Brown, 1987; Resing, 2000; Resing & Elliott, 2011). Deze prompts worden gradueel aangeboden, wat wil zeggen dat de prompts op hiërarchische wijze worden gegeven waarbij de prompts steeds specifieker worden, van algemene metacognitieve prompts, naar steeds specifiekere cognitieve prompts met als allerlaatste stap *modelling*, het voordoen van de juiste oplossing. Doordat de prompts op deze manier worden aangeboden, is het mogelijk om een inzicht te verkrijgen in de instructiebehoeften van een kind: heeft een kind bijvoorbeeld voornamelijk baat bij algemene metacognitieve instructie, of heeft dit kind specifiekere, op de taak afgestemde, cognitieve instructie nodig.

De dynamische tests die in de studies van deze dissertatie zijn gebruikt bestonden uit het oplossen van geometrische analogieën van het type A:B::C:?. Aangenomen wordt dat analogisch redeneren, wat een subtype is van inductief redeneren, een belangrijke rol speelt in het alledaags leren van kinderen (Richland, Morrison, & Holyoak, 2006). Prestaties op analogische redeneertaken zijn dan ook gerelateerd aan schoolprestaties (Balboni, Naglieri, & Cubelli, 2010) en individuele verschillen in IQ-scores en fluïde intelligentie (Caropreso & White, 1994; Vendetti, Wu, & Holyoak, 2014).

Het concept hoogbegaafdheid werd ook nader beschreven in Hoofdstuk 1. Waar in het begin van de 20e eeuw hoogbegaafdheid voornamelijk gelijk gesteld werd aan een hoog IQ (Terman, 1925), wordt in recentere definities aangenomen dat hoogbegaafdheid een dimensioneel concept is waarbij bovengemiddelde cognitieve capaciteiten slechts een onderdeel zijn en andere factoren, zoals creativiteit en taakvolharding (Renzulli, 2005) ook een belangrijke rol spelen. Recente definities van hoogbegaafdheid houden daarnaast ook rekening met interactie met de (socioculturele) omgeving (Barab & Plucker, 2002) en het idee dat hoogbegaafdheid geen statische, aangeboren, eigenschap is, maar zich ontwikkelt gedurende een mensenleven (Subotnik, Olzewski-Kubilius, & Worrell, 2012). In de studies in deze dissertatie zijn kinderen geïdentificeerd als hoogbegaafd op basis van leerkracht- en oudernominaties, zoals veelal in de praktijk plaatsvindt (Kornmann, Zettler, Kammerer, Gerjets, & Trautwein, 2015; Threlfall & Hargreaves, 2008). Alle hoogbegaafde deelnemers aan de studies in dit onderzoek genoten daarnaast onderwijs voor hoogbegaafde en/of getalenteerde leerlingen. De kinderen uit Hoofdstuk 3, 4 en 5 hadden bovendien een percentielscore van tenminste 90% op de Raven Standard Progressive Matrices Test, een test die de fluïde intelligentie meet. Dit houdt in dat zij voor deze test vielen onder de 10% best presterende kinderen vergeleken met hun leeftijdsgenoten.

Ten slotte werd in dit hoofdstuk een alternatieve conceptualisering van hoogbegaafdheid aangeboden, gebaseerd op het Model van de Zich Ontwikkelende Expertise van Sternberg (2001; Sternberg, Jarvin, & Grigorenko, 2011), waarin hoogbegaafdheid wordt gezien als een vorm van zich ontwikkelende expertise. Volgens dit model spelen vijf factoren een belangrijke rol in de ontwikkeling van beginner naar expert: metacognitie, motivatie, kennis, denken en het leren zelf. Deze factoren zijn allemaal interactief – wat betekent dat zij elkaar zowel direct als indirect beïnvloeden. Middels oefening werkt een beginner naar het niveau van een expert toe. Het is hiervoor noodzakelijk dat alle factoren interacteren, waarbij motivatie als drijvende kracht functioneert en de context van het leren ook een grote rol speelt. Volgens Sternberg (2001) ontwikkelen hoogbegaafde kinderen expertise op een sneller tempo, een hoger en/of een kwalitatief verschillend niveau dan gemiddeld-begaafden.

In de studies van deze dissertatie werd onderzocht op welke wijze een aantal van deze factoren, twee aspecten van het executief functioneren, metacognitie en cognitieve flexibiliteit (in Hoofdstuk 4) een rol spelen bij het ontvouwen van expertise van zowel hoogbegaafde als gemiddeld-begaafde kinderen. Cognitieve flexibiliteit betreft het kunnen switchen tussen perspectieven en het flexibel kunnen aanpassen van het denken als de omstandigheden hierom vragen (Diamond, 2013). Deze vaardigheid wordt gezien als een belangrijk aspect van het intellectueel en cognitief functioneren. Metacognitie bestaat uit zelf-reflectieve cognitieve processen die belangrijk zijn bij het reguleren en structureren van het leerproces (Schneider, 2010). Ook werd onderzocht of testangst bij deze kinderen het ontvouwen van expertise bemoeilijkt (in Hoofdstuk 3) en wat de rol is van expertise bij transfer, het toepassen van geleerde kennis en expertise in een andere context (Hoofdstuk 5).

In **Hoofdstuk 2** werd onderzocht of 5- tot 8-jarige hoogbegaafde en gemiddeld-begaafde kinderen verschillen lieten zien met betrekking tot hun leerpotentieel en hun instructiebehoefte. Er werd gevonden dat zowel kinderen die dynamisch getest waren als diegenen die alleen herhaald oefenden vooruitgang lieten zien in analogisch redeneren De kinderen die dynamisch waren getest lieten echter meer vooruitgang zien in het aantal correct opgeloste analogieën. Daarnaast werd gevonden dat de hoogbegaafde kinderen vanaf de pre-test een voorsprong hadden op de gemiddeld-begaafde kinderen voor wat betreft het aantal correct opgeloste opgaven en deze vooruitgang behielden op de post-test. Het maakte daarbij niet uit of zij wel of niet getraind waren. De mate waarin deze kinderen verbetering in het aantal juiste opgaven lieten zien was echter gelijk aan de mate van vooruitgang van de gemiddeld-begaafde kinderen. Ook wees deze studie uit dat de hoogbegaafde kinderen een gelijke behoefte aan instructie hadden tijdens de dynamische training: zowel voor wat betreft het totale aantal als het type prompts dat zij kregen. Bovendien werd gevonden dat jongere kinderen minder vooruitgang lieten zien in het correct oplossen van analogieën en dat zij ook meer prompts nodig hadden tijdens de training dan oudere kinderen.

In **Hoofdstuk 3** stond centraal of 7- en 8-jarige hoogbegaafde en gemiddeld-begaafde kinderen verschillen lieten zien in vooruitgang in het correct oplossen van analogieën na een training of na herhaald oefenen en of testangst bij deze twee groepen kinderen een andere relatie had met hun testscores. Wederom lieten de resultaten van deze studie zien dat dynamisch testen leidde tot meer vooruitgang in het oplossen van analogieën dan herhaald oefenen. Ook lieten de hoogbegaafde kinderen, net als in Hoofdstuk 2, een voorsprong zien in het aantal correct opgeloste analogieën vanaf de pre-test tot de post-test. Zij gingen ook in dezelfde mate vooruit als hun gemiddeld-begaafde leeftijdsgenoten. Daarnaast werd gevonden dat testangst gerelateerd was aan de mate waarin kinderen vooruitgang lieten zien van sessie naar sessie. In het bijzonder werd gevonden dat de kinderen met hogere testangstniveaus meer verbetering in analogisch redeneren lieten zien na training dan hun getrainde leeftijdsgenoten met lagere testangstniveaus. Er was geen verschil in de invloed van testangst op testscores en vooruitgang in testscores tussen hoogbegaafde en gemiddeld-begaafde kinderen.

In **Hoofdstuk 4** werd onderzocht onder 7- en 8-jarige hoogbegaafde en gemiddeld-begaafde kinderen of twee aspecten van het executief functioneren, cognitieve flexibiliteit en metacognitie, gerelateerd waren aan progressie in analogisch redeneren na herhaald oefenen of een dynamische training. Er werd in dit onderzoek niet gekeken naar het aantal correct opgeloste analogietaken, maar naar het aantal correct toegepaste transformaties bij het oplossen van analogieën. Daarnaast werd ook onderzocht of de instructiebehoefte van de hoogbegaafde kinderen verschilde van de gemiddeld-begaafde kinderen. De resultaten van dit onderzoek wezen uit dat dynamisch testen leidde tot meer verbetering in het aantal correct toegepaste transformaties dan herhaald oefenen en dat de hoogbegaafde kinderen ook hier een voorsprong hadden in het aantal juist toegepaste transformaties, deze voorsprong ook behielden, maar niet in grotere mate vooruitgingen dan hun gemiddeld-begaafde leeftijdsgenoten.

Ook werd gevonden dat cognitieve flexibiliteit en metacognitie niet gerelateerd waren aan de vooruitgang van kinderen in het aantal correct toegepaste transformaties en dat kinderen met verschillende niveaus van flexibiliteit en metacognitie in gelijke mate vooruitgang lieten zien. Daarnaast was alleen metacognitie gerelateerd aan de mate waarin kinderen profijt hadden van training, waarbij de kinderen met een lagere metacognitie meer vooruitgang lieten zien na training dan hun leeftijdsgenoten met een hogere metacognitie. Er was geen verschil in de invloed van metacognitie of cognitieve flexibiliteit op de testscores van hoogbegaafde en gemiddeld-begaafde kinderen en er was ook geen verschil in de hoeveelheid instructie die beide groepen kinderen nodig hadden tijdens de dynamische training.

Het onderzoek beschreven in **Hoofdstuk 5** onder 9- en 10-jarige hoogbegaafde en gemiddeld begaafde kinderen had twee doelen. Allereerst werd onderzocht of deze twee groepen kinderen verschillen lieten zien in de mate waarin zij vooruitgang boekten in het oplossen van analogieën na herhaald oefenen of een dynamische training. Ook werd gekeken of deze twee groepen kinderen significante verschillen lieten zien in de transfer van analogische probleemoplossingsvaardigheden van het oplossen van analogieën naar het zelf construeren van een analogie tijdens een transfer-taak. Hierbij werden zowel het aantal correct geconstrueerde analogieën als de moeilijkheidsgraad van deze analogieën in ogenschouw genomen. Er werd daarbij onderzocht of het trainen van kinderen in analogische probleemvaardigheden tot meer transfer zou leiden en of de prestaties van kinderen in het oplossen van analogieën een voorspellende waarde had voor het aantal correct door hen geconstrueerde analogieën en de moeilijkheidsgraad van deze zelf-geconstrueerde analogieën.

Er werd in dit onderzoek gevonden dat dynamisch testen leidde tot meer verbetering in het aantal correct opgeloste analogieën dan herhaald oefenen. Zowel herhaald oefenen als de dynamische training leidden, in gelijke mate, tot een verlaging van de tijd die de kinderen nodig hadden om alle items van een testsessie op te lossen. Hoogbegaafde kinderen hadden ook hier een voorsprong in het aantal correct opgeloste analogieën en behielden deze voorsprong na een training of herhaald oefenen. De twee groepen kinderen hadden echter evenveel tijd nodig om de items van iedere testsessie op te lossen en verschilden ook niet in de mate waarin zij vooruitgang lieten zien in het aantal opgeloste analogieën. Dit gold ook voor de mate waarin zij per testsessie minder tijd nodig hadden om alle analogieën op te lossen.

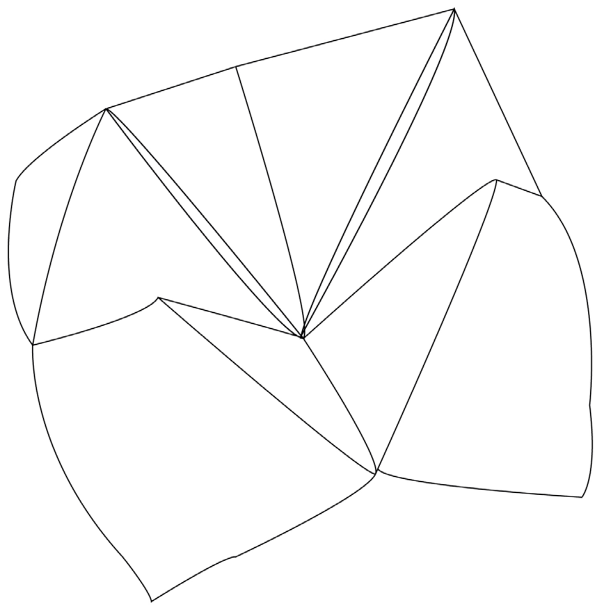
Op het gebied van transfer werd gevonden dat training geen voorspellende waarde had voor zowel het aantal correct geconstrueerde analogieën als de moeilijkheidsgraad van deze zelf-geconstrueerde analogieën. Er werden daarnaast ook geen verschillen gevonden tussen de hoogbegaafde kinderen en hun gemiddeld-begaafde leeftijdsgenoten voor wat betreft het aantal correcte zelf-geconstrueerde analogieën en de moeilijkheidsgraad hiervan. Toen de zelf-geconstrueerde analogieën van de kinderen ingedeeld werden in drie moeilijkheidsgraden, makkelijk, gemiddeld en moeilijk, bleek dat er meer ongetrainde kinderen waren die makkelijke analogieën hadden geconstrueerd dan getrainde kinderen en meer getrainde kinderen die moeilijke analogieën hadden geconstrueerd dan ongetrainde kinderen.

In **Hoofdstuk 6** zijn de uitkomsten van de studies in deze dissertatie samengevat en de hieruit voortvloeiende theoretische en praktische implicaties besproken. Allereerst kan geconcludeerd worden dat dynamisch testen van hoogbegaafde kinderen een waardevolle bijdrage levert aan het meten van hun cognitieve vaardigheden. Er werd meerdere malen gevonden dat als deze kinderen dynamisch in plaats van statisch werden getest, zij vooruitgang lieten zien, welke bevinding leidde tot de aanname dat de cognitieve vaardigheden van (hoogbegaafde) kinderen geen stabiele entiteiten zijn, maar dynamisch, die zich onder de juiste omstandigheden nog verder kunnen ontwikkelen. Het leerpotentieel van hoogbegaafde kinderen lijkt dan ook kunnen worden gezien als zich ontwikkelende expertise, zoals gesuggereerd door Sternberg (2001; Sternberg et al., 2011). Er wordt op basis van deze dissertatie dan ook aanbevolen om in het kader van predictieve vraagstukken gericht op onderwijskeuzes zich niet alleen te richten op de uitslag van een dergelijke test, maar ook op hoe het kind tot een bepaalde testscore is gekomen en welke instructie(s) hij of zij daarbij nodig had. Om de cognitieve en intellectuele vaardigheden van kinderen te meten wordt dan ook aangeraden dynamisch in plaats van statisch te testen, in het bijzonder als vermoed wordt dat een kind lage metacognitieve vaardigheden heeft of last heeft van testangst. Er kunnen verschillende factoren zijn die zorgen voor een testscore waaruit niet het cognitieve potentieel van een individueel kind blijkt, wat, als gekozen wordt voor een onderwijsvorm die niet aansluit bij de resultaten van een statische test, kan leiden tot verlies van cognitief potentieel. Daarbij is het van belang dat onderwijsprofessionals zich ervan bewust zijn dat hoogbegaafde kinderen hun cognitieve potentieel niet altijd ten volle benutten, met name bij het maken van een statische test.

Daarnaast biedt het gebruik van deze testen inzicht in de instructiebehoefte van deze kinderen. Deze bleek in de besproken studies gelijk te zijn aan de instructiebehoefte van gemiddeld-begaafde kinderen, zowel voor wat betreft de hoeveelheid als de type instructie die zij behoefden. Ook werd gevonden dat individuele hoogbegaafde kinderen, net als andere kinderen, significante verschillen laten in hoeveel zij zich verder ontwikkelen en welke instructie zij hiervoor nodig hebben, alsmede in de mate waarin zij hun opgedane kennis in een andere context kunnen toepassen. Een aanbeveling uit deze dissertatie is dan ook dat onderwijs voor hoogbegaafde en getalenteerde leerlingen voldoende mogelijkheden moet hebben tot differentiatie en aangepaste instructie, om zo tegemoet te komen aan de verschillende onderwijsbehoeften van deze groep leerlingen. Ook begaafde kinderen kunnen bijvoorbeeld behoefte hebben aan een verlengde instructie.

Bovendien laten de resultaten van de studies die onderdeel zijn van deze dissertatie zien dat hoogbegaafdheid niet een statische, stabiele categorie is, waarbij een individu simpelweg "wel" of "niet" hoogbegaafd is (Pfeiffer, 2011), maar eerder een dynamische eigenschap die zich, afhankelijk van verschillende factoren, wel of niet volledig ontwikkelt. De laatste aanbeveling van deze dissertatie is dan ook dat onderwijs aan alle kinderen, ongeacht hun begaafdheid, gebaseerd zou moeten zijn op principes van het Model van de zich Ontwikkelende Expertise (Sternberg, 1999, 2001; Sternberg et al., 2011) en het Raamwerk voor Talentontwikkeling (Olszewski-Kubilius & Thomson, 2015). Het doel van het laatstgenoemde raamwerk is het bewerkstelligen van onderwijs dat gericht is op het vormen van psychosociale vaardigheden die buitengewone prestaties, volharding en creativiteit bevorderen, opdat het beter in staat is aan te sluiten bij de onderwijsbehoeften van een diverse populatie van kinderen die de potentie hebben om te excelleren.

CURRICULUM VITAE



Curriculum Vitae

Bart Vogelaar was born on 12 August 1987 in Krimpen aan den IJssel. He obtained his secondary school *Gymnasium* diploma in 2004 from Krimpenerwaard College in Krimpen aan den IJssel. Then, he studied English, spent an exchange year studying at the University of Hull (UK), and obtained his Bachelor's degree in English Language and Literature in 2007 at Leiden University, and his Master's degree in English Language at the University of Manchester (UK) in 2008. While working full-time in different types of education, he obtained a Bachelor's degree in Psychology in 2012, and a Master's degree in Child and Adolescent Psychology in 2014, both at Leiden University. During these two studies, he met Professor Wilma Resing, who sparked his interest in dynamic testing and learning potential, and encouraged him to undertake a PhD project into the dynamic testing of gifted children as an external PhD candidate. While working on his PhD dissertation under the supervision of Professor Wilma Resing and dr. Lianne Hoogeveen, he continued to work full-time in education, both as a teacher educator at the Amsterdam University of Applied Sciences, and as a lecturer at the Developmental and Educational Psychology Unit of the Department of Psychology at Leiden University. Currently, he is still working as a teacher educator at the Amsterdam University of Applied Sciences, teaching courses on adolescent psychology and coaching, and as a lecturer at Leiden University, where he is involved in teaching several bachelor's and master's courses, as well as internship and thesis supervision.