Progression paths in children's problem solving:

The influence of dynamic testing, initial variability and working memory

Wilma C.M. Resing[1], Merel Bakker[1], Christine M.E. Pronk[1], & Julian G. Elliott[2]

[1]Developmental and Educational Psychology

Leiden University, the Netherlands

[2]School of Education

Durham University, UK

Personal Uncorrected Copy

JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY, 153 (2017) 83–109

http://dx.doi.org/10.1016/j.jecp.2016.09.004

ONLINE: 13 OCTOBER 2016

Progression paths in children's problem solving:

The influence of dynamic testing, initial variability and working-memory

Abstract

The current study investigated developmental trajectories of analogical reasoning performance of 104 7-to-8-year-old children. We employed a microgenetic research method and multilevel analysis to examine the influence of several background variables and experimental treatment on the children's developmental trajectories. Our participants were divided into two treatment groups: repeated practice alone and repeated practice with training. Each child received an initial working-memory assessment, and was subsequently asked to solve figural analogies on each of several sessions. We examined children's analogical problem-solving behavior and their subsequent verbal accounts of their employed solving processes. We also investigated the influence of verbal and visual-spatial working-memory capacity and initial variability in strategy-use on analogical reasoning development. Results indicated that children in both treatment groups improved but gains were greater for those who had received training. Training also reduced the influence of children's initial variability in the use of analogical strategies with the degree of improvement in reasoning largely unrelated to working memory capacity. Findings from this study demonstrate the value of a microgenetic research method and the use of multilevel analysis to examine inter- and intra-individual change in problem-solving processes.

Keywords Dynamic testing; Microgenetic; Figural analogies; Inductive reasoning; Working memory; Multilevel analysis

Progression paths in children's problem solving:

The influence of dynamic testing, initial variability and working-memory

Introduction

Fine-grained investigation of children's cognitive abilities, including the influence of training, is complex, and children's performance on reasoning and problem solving tasks has proven significantly more variable over time than many researchers and practitioners had assumed (Bjorklund & Rosenblum, 2001; Siegler, 2007). Substantial variability in solving cognitive tasks, between and within participants, has been demonstrated in a variety of domains, such as math, spelling and problem solving (Siegler, 1996). Studying this variability could provide us with more insights in how children determine the best ways of solving tasks. This approach may also help us gain a greater understanding of individual differences in the development of strategy-use over time and with increasing experience (Siegler, 2006).

Our study sought to examine variability in children's strategic behavior in analogy solving when receiving either a series of unguided practice sessions alone or this in combination with a training procedure derived from dynamic testing. We also studied the possible influence of working memory differences on analogy performance, as a differential relationship between working memory and analogical reasoning has been found for children trained in dynamic testing settings (e.g., Stevenson, Bergwerff, Heiser, & Resing, 2014; Swanson, 1994; 2011). The main focus of our study was therefore on 'the rate of change' and 'variability' dimensions of Siegler's (1996) 'overlapping waves' theory.

Inductive reasoning tasks, such as classification, analogies or series completion, all require comparable underlying problem-solving processes: starting with specific observations of the task details under consideration, and identification of a rule that leads to the solution must be detected

and formulated. This rule finding process can be reached by means of systematic comparison processes, which involve finding similarities and/or differences between task attributes and/or relations among attributes (Holyoak & Nisbett, 1988; Klauer & Phye, 2008; Perret, 2015). Analogical reasoning involves a basic inductive process that plays an important role in a number of higher cognitive processes (Halford, 1993; Morrison et al., 2004; Richland & Simms, 2015). In analogical studies, base items (for example, white: black) and targets (for example, snow: ??) domain have to be compared in order to find and formulate a relational correspondence existing between them (e.g., Holyoak, 2012, Thibaut & French, 2016).

The development of analogical reasoning in children and its role in instruction and classroom learning have been the focus of much research (Csapó, 1997; Goswami, 2002; Klauer & Phye, 2008; Kolodner, 1997; Richland, Morrison, & Holyoak, 2006; Richland & Simms, 2015; Singer-Freeman, 2005; Vosniadou, 1989). The development and training of children's ability to reason by analogy have also been studied extensively (Alexander et al., 1989; Alexander, Willson, White, & Fuqua, 1987; Goswami, 2013). In most studies, children older than 6 years have displayed clear improvements in analogical reasoning after receiving a (brief) period of training or, alternatively, after having been given extensive instructions or training for, for example, verbal analogies (Resing, 2000), physical problem analogies (Tunteler & Resing, 2007), concrete pictorial analogies (Hessels-Schlatter, 2002;Stevenson, Resing, & Froma, 2009), and classic geometric analogies (Hosenfeld, Van der Maas, & Van den Boom, 1997; Tunteler, Pronk, & Resing, 2008). In contrast, younger children have tended only to show such gains when they had received extensive training (Alexander et al., 1989; Tunteler & Resing, 2007).

Studies focusing on the development of inductive (analogical) reasoning have often utilized cross-sectional designs (e.g., Chen, 1996; Chen & Daehler, 1989; Richland et al., 2006; Singer-Freeman, 2005; Thibout, French, & Vezneva, 2010). However, these are likely to provide

an incomplete picture of the dynamics of variability and change (e.g., Granott, 2002; Kuhn, 1995, 2002; Siegler, 2006). Very few studies have compared changes in analogical reasoning over time which have been induced by repeated practice and/or training (e.g., Alexanderet al., 1989; Hosenfeld et al., 1997), a shortfall which the current study sought to address.

One valuable approach to investigating learning trajectories is that which employs a microgenetic research design (Siegler, 2006; Winne & Nesbit, 2010). Such designs are characterized by frequently repeated (trial-by-trial) assessment sessions given within a relatively short period of time. They typically utilize somewhat basic instructions or unguided practice sessions, and typically yield highly frequent of observations relative to the rate of change. Hence, changes in reasoning can become visible closely to the moment they happen, thus enabling the discovery of natural developmental and learning trajectories. These trajectories may be considered natural if the practice sessions include no explicit forms of training, such as the provision of elaborate instructions or prompting (Flynn & Siegler, 2007; Siegler, 2006; Siegler & Crowley, 1991). In the present study we investigated whether problem solving trajectories involving analogy tasks showed differing pathways when these were acquired through more 'natural' unprompted opportunities than when a short training procedure is included.

Dynamic testing, using a test-training-test format, has become increasingly used for the study of inductive reasoning (Bethge, Carlson, & Wiedl, 1982; Hessels-Schlatter, 2002; Resing & Elliott, 2011; Tzuriel & Flor-Maduel, 2010). Key to this approach is the incorporation of feedback and training during the testing phases (Elliott, Grigorenko, & Resing, 2010; Sternberg & Grigorenko, 2002). Conventional, static tests are considered to be a satisfactory means of assessing already developed abilities. In contrast, dynamic measures are designed to assess developing or yet-to-develop abilities which are the products of underlying, but often unrecognized, cognitive capacities (Elliott et al., 2010; Haywood & Lidz, 2007; Sternberg &

Grigorenko, 2006). Dynamic testing has been found to yield insights into strategies used by the examinee and their responsiveness to examiner assistance and support. In line with the pioneering studies of Campione, Brown, Ferrara, Jones & Steinberg (1985), we sought to examine children's potential for learning in the domain of analogical reasoning. Our primary focus was on the number of prompts needed to solve the analogy items. Here, the elements of the task that cannot be accomplished unassisted are controlled, with graduated standardized support given whenever the child is unable to make unaided progress.

In the current study we combined both dynamic and microgenetic methods of studying children's development and learning. We examined whether a dynamic testing approach in combination with unguided practice, versus unguided practice alone, would result in different developmental trajectories in young children's solving of analogy tasks.

The first few years of primary school are considered to be a period of rapid development of analogical reasoning ability (Siegler & Svetina, 2002), a time when high levels of intra-individual variability in solving analogies are likely to be found. Siegler (2007), for example, posits that cognitive variability is an important variable in understanding, predicting, and describing the amount and type of cognitive change. He refers to cognitive variability as the differences between children in terms of change agents, growth trajectory, generalization, and speed of change, but also changes within the individual child's repertoire of strategies. As a way to accurately assess variable strategy-use, Siegler (2007) stresses the value of trial-by-trial assessments focusing on the acquisition of new strategies, an increased usage of the most advanced strategies, increasingly efficient execution of strategies, and improved choices among strategies. In the current study, we employed a microgenetic, trial-by-trial assessment in order to investigate variability in children's use of strategies when solving analogies and gauge their subsequent progress in their use of problem-solving strategies.

In a microgenetic study of matrix completion, Siegler and Svetina (2002) found that the performance of 6-8 year old children improved considerably on a multiple choice matrix task as a result of repeated practice experiences. After discovery of the correct way of solving the tasks, children's use of the improved strategy rapidly became dominant. However, patterns in strategy-use may differ when items require the construction of a solution, rather than selection using a multiple-choice format. Tunteler et al. (2008) conducted a microgenetic study with 6-7-year-old children, using open-ended (constructed response) geometric analogies, behavioral and verbal outcome measures, and minimal instruction. Their results indicated that improvement in analogical reasoning often consisted of progression from incomplete to complete analogical answers. Children appeared to possess some rudimentary form of analogical reasoning skill that was accelerated by the opportunity to practice. After training, however, the children were largely able to explain their use of correct analogical strategies in solving the tasks. Further investigation revealed several subgroups of children that varied in the extent to which they could provide oral explanations of their approach to solving the problems (Tunteler et al., 2008). We sought to build on these findings in the current study by investigating both inter- and intra-individual paths of change by examining a) children's behavioral strategies and b) their subsequent accounts of the procedures they had employed to tackle the problems.

An accumulating body of evidence suggests that working-memory capacity has a bearing upon the ability to solve complex reasoning tasks (Halford, Wilson, & Philips, 2010; Morrison et al., 2004; Primi, 2001). Many studies have explored the manner and extent to which inductive reasoning is related to working-memory capacity (Arendasy & Sommer, 2005; Engle, Tuholski, Laughlin, & Conway, 1999; Meo, Roberts, & Marucci, 2007; Richland et al., 2006; Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004; Waltz, Lau, Grewal, & Holyoak, 2000). When solving analogies, children's working memory appears to be particularly important for encoding

and processing the terms of the analogy (Sternberg & Rifkin, 1979). The type of relationship or task that needs to be managed appears to be influenced by the differential involvement of separate components of working memory. Various components have been investigated in a variety of inductive reasoning or academic tasks in different age groups (Alloway & Passolunghi, 2011; Raghubar, Barnes, & Hecht, 2010). Age-related, differential involvement of these components in different types of tasks, were, among others, demonstrated by Alloway, Gathercole and Pickering (2006). They found that children as young as 4 years exhibit a structural organization of memory into both a domain-general component for processing information and verbal and visual-spatial domain-specific components for storage. In the present study, we focused on the potentially differential involvement of verbal and visual-spatial working-memory components in analogical reasoning development. This was a consequence of earlier research in dynamic testing which has reported that both visual-spatial and verbal reasoning components play a role in solving visual-spatial analogies. This is particularly apparent when, as part of the assessment, children are asked to explain their problem solving procedures (Stevenson, Heiser, & Resing, 2013; Tunteler et al., 2008). In the present study we explicitly focused on the differential involvement of verbal and visual-spatial working-memory components, to examine their possible role in respect of changes in analogical reasoning over time. We thought it important to examine these components separately with a working-memory assessment that made sufficient storage and processing demands (Alloway, 2007) and which would help us explore their influence on analogical reasoning (Resing, Xenidou-Dervou, Steijn & Elliott, 2012).

Multi-level growth-curve analysis (e.g., Hox, 2010; see Method) enabled us to model the average growth trajectories of both groups of children and each child individually (Hox, 2010). Variability in strategy-use would be an important variable in predicting growth trajectories (see

also Siegler, 2007). We investigated systematic variation between these trajectories as a function of our background variables – the verbal and spatial working-memory components – and the experimental treatment [i.e. the dynamic test training procedure (Van der Leeden, 1998)]. In addition, our study also included initial variability in strategy-use as a potentially interesting background variable.

In sum, the aim of the current study was to model the influence over time of a) repeated unguided practice, b) repeated unguided practice plus training derived from a dynamic testing procedure, c) verbal and visual-spatial working memory, and d) initial variability in problem-solving procedures, on the problem-solving processes and individual outcome trajectories of young children. Using figural analogies tasks, we examined whether unguided practice with or without dynamic training would result in different changes in strategy-use and in performance accuracy. In addition, we anticipated that measures of working memory and initial variability in children's problem solving would provide us with a richer understanding of factors underpinning individual growth trajectories in analogical reasoning.

Firstly, we examined whether grade 2 children changed their approach to tackling figural analogy tasks on the basis of repeated unguided practice and having received, or not received, the dynamic testing intervention. The specific components measured were the children's accuracy on the task, the use of various task-solution components as a measure of the (in)completeness of children's performance on these tasks and their verbal explanations of their activity as a measure of the quality of their solving strategies (Tunteler & Resing, 2010). We hypothesized that a) repeated practice would lead children to independently adopt more advanced solution methods. However, we further hypothesized that b) trained children would make greater gains than those who received no such assistance (Resing et al., 2012; Siegler & Svetina, 2002).

Secondly, it was expected that spatial working-memory capacity would be related to figural analogical reasoning performance at the first session and also to improvement following repeated practice. Our reasoning for this was that children have to remember visuo-spatial information, for example, the position, size, color, and orientation of the various attributes of both the base and the target domains of the analogy in order to construct relational representations (Logie, Gilhooly, & Wynn, 1994; Rasmussen & Bisanz, 2005, Tunteler et al., 2008). Children with smaller spatial working-memory capacity were expected to make limited progress as a result of repeated practice as their workspace for constructing relational representations is assumed to be more limited (Halford et al., 2010). Those who had received training, however, were expected to subsequently demonstrate a higher rate of progress (Carr & Schneider, 1991; Halford et al., 2010; Morrison, Doumas, & Richland, 2011). The rationale behind this hypothesis was that dynamic testing was expected to alleviate any working-memory limitation by breaking down the analogical reasoning process into small steps that can be processed serially and by providing relational knowledge (Halford et al., 2010; Morrison et al., 2011). This might help these children to catch up with peers with superior working-memory capacity. Children with larger spatial working-memory capacity, on the other hand, were expected to show a more gradual pattern and rate of change over time when receiving repeated practice alone, but nevertheless gain additional benefit from the training procedure (Tunteler & Resing, 2010).

Thirdly, it was expected that the relationship between verbal working-memory capacity and performance on the figural analogies test would become less strong, both over time and after training on the grounds that our dynamic training procedure would reduce cognitive load (Alloway & Gathercole, 2006; Haavisto & Lehto, 2004; Halford et al., 1998, 2010; St. Claire-Thompson & Gathercole, 2006).

Fourthly, we anticipated that variability in an individual's strategy-use at the first session would be positively related to rate and amount of change in solving analogies, especially for the unguided practice sessions (Siegler, 2007). However, the progress of children receiving training in addition to repeated practice was expected to be less related to their initial performance as the assistance was intended to tap into underlying variability in potential rather than current, unassisted performance (Grigorenko, 2009). We therefore predicted that those children who displayed limited analogical reasoning before training would improve more rapidly than their peers, whereas higher performers before training would show a more gradual increase in the quality of their reasoning (Tunteler et al., 2008; Tunteler & Resing, 2007).

Finally**,** the paths of change in children's solving procedures were explored through inspection of the various group progression lines regarding the number of correctly solved analogies, the use of transformations, and their explanations of the approaches they adopted. We visually inspected how children [differentially] changed their solving procedures over sessions towards the use of more sophisticated problem-solving procedures (Siegler, 2006). The influence of children's level of working memory, variable analogical solving skills at the start of the experiment, and training on this progression in solving procedures was further explored.

Method

Participants

One hundred-and-four children (51 boys and 53 girls) with a mean age of 7.8 years (age range 7.0–8.7 years) participated in the study. The children were recruited from eight primary schools, located in midsized towns in the Netherlands. The schools were selected on the basis of their willingness to participate in what was clearly a time-consuming exercise. The children attended classes at second grade level and represented a full range of socio-economic backgrounds. Parental consent for participation was obtained in all cases. Dutch was the first language spoken

at school and at home for all participants. Data collection was undertaken by trained postgraduate students with experience of educational testing. All children took full part in the relevant testing sessions.

Design

The study employed a pre-test-post-test-2-experimental-groups randomized block design, with blocking based on the Exclusion subtest from a Dutch child intelligence test, the RAKIT (Bleichrodt, Drenth, Zaal, & Resing, 1984). On the basis of the blocking procedure, children in each school were, randomly allocated to one of two treatment groups: an 'unguided practice' group; or an 'unguided practice plus training' group, subsequently identified in this paper as the 'training group'. During the first week of the study, each child was given an inductive reasoning test (Exclusion) and measures of spatial and verbal working memory. Subsequently, a microgenetic design with two pre-tests and two post-tests was employed. Children in the training group received training between the pre-test and post-test sessions. In all other respects they received the same inputs as the children in the unguided practice group. Thus, other than the training session, both groups tackled the same analogy items without other forms of training, instruction, help or feedback. Unguided practice sessions ranged from 20-40 minutes per child and were of equal duration for both groups of children. The training session took 30-60 minutes per child. Table 1 provides an overview of the experimental design used in this study.

Insert Table 1

Instruments

Exclusion. The Exclusion subtest from a Dutch child intelligence test (RAKIT: Bleichrodt et al., 1984), served as the means for blocking the groups. This measure taps the

child's ability to detect rules by means of induction, an important prerequisite for successful inductive/analogical reasoning. The subtest (α=.87) consists of 40 items each containing four abstract figures. Three of the figures can be grouped together according to a rule, and the child's task is to select the figure that does not conform to this.

Verbal and spatial recall. The screening measure from the computerized Automated Working Memory Assessment (AWMA) battery (Alloway, 2007), administered during the pre-testing phase, was used to measure verbal and visual spatial working-memory capacity. The AWMA assesses both the simultaneous storage and processing of information. The screening version exists of two working-memory subtests: Listening recall (verbal WM; test-retest reliability r=.88) and Spatial recall (visuo-spatial WM; test-retest reliability r=.79)). The listening recall subtest utilizes sequences of spoken sentences. The child must indicate whether the spoken sentence is true or false. At the end of each item, the child has to recall the final word of each sentence, in the order in which the sentences were presented. In the spatial recall subtest, the child sees a picture with two abstract shapes; the shape at the right has a red dot somewhere. The shape might be rotated or flipped. They are then required to indicate whether the shapes to the left and right are the same or not. At the end of each item, the correct location of the red dot must be identified (Alloway, 2007).

Animalogica: Dynamic test of Figural reasoning

Figural analogies: pre- and post-tests. The analogical reasoning tasks utilized in this study consisted of an adapted open-ended response version of the concrete figural test *Animalogica* (see, Stevenson et al., 2009; Stevenson, Touw, & Resing, 2011; Stevenson, Heiser, & Resing, 2016). To solve the analogies (A:B = C: ?, see Figure 1), the children were required to encode item attributes and infer a rule between two given animal pictures (A, B [the base]) and then apply this to the third picture (C [the target]) of the analogy. Finally, a fourth animal picture

had to be selected (from a range of possibilities), so that the relationship between the third and fourth picture equaled that of the first two pictures (e.g., Sternberg & Rifkin, 1979). This open-ended response version enables children to identify the correct rule and then search for the right answer pictured on one of 36 cards, variously showing six types of animals, three colors, and two sizes. The cards were printed on both sides with the same images, although facing in opposite directions, enabling the child to transform both quantity and orientation of the animals for solving the analogy. Four parallel sets of 20 items each, increasing in difficulty level, were used. The sets were designed to look different to the children but the changes in animal type and color were of a similar level of difficulty.

Insert Figure 1

The items of each set varied in difficulty level from two to six transformations involving, size, color, quantity, orientation, position, and type of animal. In the analogy in Figure 1at the left, for example, we can detect one small yellow lion (A) that has been transformed into two small yellow lions (B). We also see two small blue horses looking into the same direction as the lions (C), so we need to look for one small blue horse looking to the right to fill D. Based on the cards representing children's responses, it was possible to see which transformations children had used to construct their answer, and if and how many of the transformations necessary to solve the items were actually used.

Figure analogies training. The training procedure used in the present investigation was originally pioneered by Campione et al., (1985), and has been successfully further developed and utilized in studies on dynamic testing (Resing, 2000; Resing & Elliott, 2011). The procedure involves the use, during training, of a series of adaptive and standardized, hierarchically ordered, prompts that proceed from general, metacognitive, to increasingly task-specific prompts. The

prompts are only provided when a child is unable to proceed independently. Thus, the children are provided with the minimum number of prompts possible to enable progression through the test. The training of the dynamic test procedure consisted of an age-adapted set of 7 concrete figural analogies similar to those employed in the other sessions (adapted from Stevenson et al., 2009, 2011). The procedures involved are described in Appendix A. Stevenson et al. (2016) reported internal consistencies for the pre-test and training procedure (α=.78, =.78) and post-tests 1 and 2 (α=.90 and .85, respectively).

Procedure

All children were, seen individually, once a week over a two month period. The same testers were employed across the full series of six sessions. During the first week of the study, the children were given inductive reasoning and spatial and verbal working memory tasks. Subsequently, two pre-tests and two post-tests were administered (see overview of design). Between these, children in the training group received 30-60 minutes of training. The length of these sessions depended, in part, upon the number of hints they required. Other than the training session, both groups tackled the same analogy items without any other forms of training, instruction, help or feedback. Unguided practice sessions ranged from 20-40 minutes per child and were of equal duration for both groups.

At the start of each session, the child was presented with a booklet containing the analogies, and baskets with small animal cards for constructing the correct answers in accordance with the transformations used in the items. The examiner showed the animal cards and explained their features: color, size, and the possibility to flip to the opposite direction. The examiner then identified the first analogy and stated that this was a ‘kind of puzzle’ with three boxes containing animals and a fourth empty box (C-term or D-term), in which the child needed to construct the solution to ‘the puzzle’ using the animal cards. After producing each solution, the child was

asked how he or she had solved ´the puzzle´. All verbal and behavioral responses were recorded by video and audio. Examiners also recorded children's behaviors and explanations on coded response forms. One author always checked and double-checked the scoring of children's solutions using these recordings.

Scoring and analyses

Scoring. Each child obtained, per practice session, a '*Complete Analogies Score'*: the sum of all analogies that were completely and accurately solved (score range 0-20); a '*Transformations Correct Score':* the sum of accurate transformations as evidenced by the child's behavioral solutions. These were calculated from data on the answer forms and findings were subsequently double-checked by means of scrutiny of the video materials (score range 0-110). The transformations refer to the changes that occur in the items in relation to size, color, quantity, orientation, position, and type of animal. We counted the number of transformations children laid down correctly, both vertically and horizontally in the matrix. In the left part of Figure 1, for example, there are three transformations: the quantity changes horizontally (one to two animals), the type of animal changes vertically (lion to horse) as does the color of the animals (blue to red). The matrix at the right of Figure 1 includes 6 transformations (horizontally: position, quantity, orientation, and type of animal, and vertically: size and color). If there were similar transformations in both columns and rows, they were counted twice. The children also obtained a *'Transformations Explained Score':* the sum of all accurate transformations that were followed by explanations as to 'how they solved the puzzle' (score range 0-110). The child had to verbalize each transformation, for example, "this one is blue also", or "here it was a horse and now it is a dog" or "here I see one dog and over there, there are two dogs". Utterances that indicated no underlying reasoning such as "just like that", with, or without pointing, were not scored, pointed directions (horizontal/vertical) were scored The Complete Analogies Score was

employed as a measure for accuracy, the number of Transformations Correct Score as a measure for the (in)completeness of the children's problem solving, and the Transformations Explained Score as a measure of the quality of their problem solving strategies.

We identified two groups of children on the basis of their problem-solving behavior at the first session: 1) children showing consistent inadequate, non-analogical reasoning (<20 percent correctly solved items), and 2) children showing variable, adequate and inadequate, reasoning (20-80 percent correctly solved items). This 20% benchmark was chosen on the basis of the mean of the scores at pre-test: the *Animalogica* tasks had a high level of difficulty. Children answered, on average, about 4 items correctly during the pre-test (see Table 2). Only one participant answered more than 80 percent of the items correctly (i.e., consistent analogical strategy-use). This child was reassigned to the group of children who showed variable strategy-use.

Analyses. Repeated measures analysis is frequently used to analyze data derived from repeated measurements of the same individuals. However, microgenetic data sets can be considered to be comprised of multilevel data, where repeated measurements are nested within individuals (Hox, 2010; Kreft & De Leeuw, 1998; Van der Leeden, 1998). Multilevel analysis (MLA) was used to analyze the data because it allowed us to model the intervention and effects of practice separately, and over time. The analyses were carried out in R (R Core Team, 2014), as a series of logical steps involving the increased addition of predictors, in line with our hypotheses.

Linear Mixed Modelling analysis, utilizing a multilevel approach (MLA applied with the lme4 package for R Statistical Software; Bates, Maechler, Bolker, & Walker, 2015), appeared to be particularly valuable for the present study as it enabled us to inspect growth trajectories based on data obtained from repeated measurement moments (Level-1), for each individual (Level-2). It also enabled us to investigate systematic variation between these trajectories as a function of our

background variables and experimental treatment (Van der Leeden, 1998). Additionally, Linear Mixed Modeling enabled us to add two types of explanatory variables to the model: time-constant and time-varying variables. This allowed us to model both the average growth trajectories of various groups of children, as well as the individual growth trajectories of each child (Hox, 2010). The raw data were used to describe the development of children's analogical reasoning performances defined in terms of accuracy, number of transformations, and number of transformations mentioned in verbalizations, over sessions (see Table 2). Then, a series of models was compared for each of the three dependent variables using maximum likelihood estimation (FML).

First, an unconditional means model (model 0) was carried out that included a random intercept representing fixed (average) and random (variance) effects to examine variation in the intercept (i.e., variable mean) of each variable. In model 1, the unconditional growth model, we included the linear effect of time. These models were carried out to analyze the variance in the three dependent variables between children and over time within children, respectively. The subsequent models included predictor variables to explain remaining variance. The following predictors were added to the model: group verbal and spatial working memory, and initial variability.

Each successive model included an additional predictor variable or interaction between predictors. The linear effects of time were included in all models except for model 0. Likelihood ratio (LR) tests (using the Chi-square distribution) and fit indices (AIC and BIC) were examined to assess how the model fit of the new model compared to that of the previous fitted model. All of the variables contained a meaningful 0-point to facilitate the interpretation of the outcomes of the analyses (Hox, 2010). For reference purposes, the regression equations for the final models are displayed in Appendix B.

After running the MLA's, we focused on more in-depth analyses of group-growth curves of analogical reasoning over time. Siegler (2007) posited the benefit of trial-by-trial assessments in microgenetic studies. To achieve this, we examined: the number of complete analogical solutions; the number of correct transformations the child produced in both incomplete and complete analogical solutions; and the number of these correct transformations for which the child was able to describe their strategy.

Results

Psychometric analyses regarding the pre-test outcomes revealed a high internal consistency for the number of complete analogical solutions (α=.90) .Test-retest coefficients were calculated for test session 1 and test session 2 (complete analogical solutions, r=.86, p<.001; transformations correct, r= .94, p<.001; transformations explained, r=.89, p<.001).

Before examining the findings from each of our research questions in detail, we checked for possible initial differences between the training and practice groups using t-tests. There were no differences between the groups in the Exclusion test *(*t(102)=−-1.14, p=.26) and in the number of complete analogical solutions (t(102)=0.39, p=.70), transformations (t(102)=0.76, p=.45), or explanations (t(102)=0.78, p=.44) at Session One. Means and standard deviations per session and group are provided in Table 2. Verbal working memory scores (t(102)=0.47, p=.64) and spatial working memory scores (t(102)=-0.50, p=.62) also did not differ between the two groups (see Table 3).

Insert Tables 2 & 3

Growth curve analyses (MLA) were based on raw scores, and were used to model growth across three outcome variables: the number of a) complete solutions; b) correct transformations

and c) explained transformations. Several hierarchical analyses were performed to find the best fitting growth model for each dependent variable separately.

Complete solutions (accuracy)

Model 0 (the unconditional means model, see Table 4) revealed a significant fixed effect of the intercept (p<.001). The intra-class correlation coefficient (ICC) indicated that 76% of the total variation in the analogy scores was attributable to inter-child differences. We included our time predictor into the level-1 sub-model in order to explain the residual within-child variance (7.98).

In model 1 (the unconditional growth model), the fixed effect of linear time was, as expected, significant. Children, on average, increased in accuracy across sessions. There was a positive covariance (.22) between the slope and intercept, which revealed that those with higher initial analogy scores generally demonstrated higher rates of growth. Inspection of the variance components revealed large remaining variance in complete solutions both between, and within, children. The R^2 value indicated that 52.2% of the within-person variation in accuracy was accounted for by the linear effect of time.

Insert Table 4

In model 2, (continuing examination of hypothesis 1), the positive fixed effect for group showed that the children generally increased their accuracy across sessions as the estimated rate of change (0.902) for repeated practice showed. The trained children obtained, as expected, higher scores than children in the unguided practice group. This difference in performance took place, in particular, between session 2 and 3. The negative time x group effect indicated that trained children's reasoning accuracy, unexpectedly, decreased from the first to second post-test (session 3 to 4; see Figure 2). The outcomes of model 2 also revealed large remaining inter- and

intra- variance children's accuracy scores. The variance in Time indicated that unexplained variance accounted for children's rate of change.

Insert Figure 2

In model 3 we inspected the effects of both verbal and spatial working memory on initial status (see hypotheses 2 and 3). An LRT indicated that the inclusion of these predictors resulted in an improved model (X^2=23.66, *df*=2, *p*<.001). As expected, spatial working memory significantly added to the prediction of the number of complete solutions at pre-test. Children with greater spatial working-memory capacity obtained higher accuracy scores at session 1. The fixed effect of verbal working memory was not significant. Model 4 added the interaction effects between time and verbal working memory and spatial working memory respectively, and this led to a significant improvement in model fit (X^2=9.01, *df*= 2, *p*=.01). The interaction of verbal but not spatial working memory with time was significant. Contrary to our expectations, spatial working memory was not related to improvement through repeated practice. The interaction between time and spatial working memory was therefore not further included in modelling. Model 5 added the interaction effects of group with verbal working memory and spatial working memory respectively. Model fit did not improve, indicating that the children were able to benefit from dynamic testing and improve their analogical reasoning performance irrespective of the capacity of their working memory. The inclusion of the three-way interactions (spatial working memory x group x time and verbal working memory x group x time) also did not improve model fit.

In model 6, initial variability was included and this led, unsurprisingly, to an improvement in model fit as indicated by smaller AIC and BIC fit indices. However, adding the

interaction between initial variability and time in model 7 did not result in a better fitting model. Contrary to our expectations, variability in children's strategy-use at the first session was not related to the rate of change in solving analogies. In model 8, we included the interaction of initial variability and group, which significantly improved model fit compared to model 6 (X^2=6.81, *df*=1, *p*<.01). The negative interaction effect revealed, as expected, that the dynamic-test training reduced the influence of children's initial variability in the use of analogical strategies. Including the three-way interaction of initial variability x group x time did not improve model fit. Results of the likelihood ratio tests and inspection of the AIC and BIC indices showed that model 8 was the best fit to our data. Model validation with residual plots (Package '*ggplot2'*; Wickham, & Chang, 2015) did not reveal any violation of assumptions.

Summary for accuracy. Our final model revealed that the training was successful in improving children's accuracy. However, the benefits of training were not fully sustained throughout the study sessions. Our findings further indicated that children with low initial variability in analogical reasoning profited more from the dynamic-test training than children who already demonstrated some analogical reasoning at pre-test. At this first session, spatial working memory was positively related to the number of complete analogical solutions, whereas (in the final model) verbal working memory was a significant predictor of children's initial status and rate of change. Children with greater verbal working-memory capacity were found to improve more in accuracy across test sessions.

Number of (explained) transformations

The outcomes of the various models we tested regarding children's use of correct transformations and the number of explained transformations are depicted in Tables 5 and 6, respectively. The unconditional means model (model 0) of both outcome variables showed a significant fixed effect of the intercept (*p*<.001). For both outcome variables, an ICC of .81 was

found, indicating that 81% of the total variation in the number of (explained) transformations was attributable to inter-child differences. The unconditional growth model (model 1) of both outcome variables revealed significant fixed effects of linear time. The R^2 values indicated that 52% and 36.7% of the within-child variation in the number of transformations and the number of explained transformations respectively, was systematically associated with linear time.

Insert Tables 5 and 6

Model 2 included group as a predictor. A significant time effect was found for the number of transformations, indicating that children, on average, increased their number of correct transformations across sessions as a result of repeated practice. Unexpectedly, however, repeated practice did not lead children to significantly increase their number of explained transformations across sessions. The significant fixed effects of group in the models of both outcome variables revealed that the trained children correctly produced and explained more transformations than those who only had unguided practice. This difference seemed to be most evident between sessions 2 and 3. However, the effect of training again decreased at the last session for both outcome variables (transformations and explained transformations), as indicated by negative time x group effects (see Figure 2).

In model 3, we included the main effects of verbal and spatial working memory. This resulted in improved models for the number of both transformations (X^2=18.87, df=2, p<.001) and explained transformations (X^2=21.18, df=2, p<.001). As expected, spatial working memory but not verbal working memory added significantly to the prediction of the number of (explained) transformations at the first session. Model 4 included the interaction effects of time with verbal working memory and spatial working memory. Model fit did not improve for the number of

transformations, or explained transformations. In model 5 the interaction effects of group with verbal working memory and spatial working memory were added. Similar to model 4, non-significant improvements were revealed, indicating that verbal and spatial working memory capacity did not influence children's training benefits. The 3-way interactions (spatial working memory x group x time and verbal working memory x group x time) failed to improve model fit for both the number of transformations and the number of explained transformations. The main effect of verbal working memory was no longer included in the models of these outcome variables after this point.

The main effect of initial variability was included in model 6, which, again, led to an improvement in model fit for both outcome variables as indicated by smaller AIC and BIC fit indices. The interaction of initial variability and time in model 7 did not result in a better fitting model for the number of transformations or explained transformations.

We included the interaction between initial variability and group in model 8, which led to a significant improvement in fit compared to model 6 for the number of transformations (X^2=16.89, *df*=1, *p*<.001) and the number of explained transformations (X^2=6.14, *df*=1, *p*=.01). The negative interaction effect indicated that dynamic-test training reduced the influence of children's initial variability in the use of analogical strategies. Including the interaction of initial variability x group x time did also not improve model fit for the number of transformations or explained transformations. Again, model 8 was selected as having the best fit to the data for both outcome variables. Model validation ('ggplot2') showed that assumptions were met.

Summary for (explained) transformations. The final models indicated that dynamic-test training was successful in improving children's number of transformations and explained transformations. However, as has been shown, the initial benefits of training were not fully maintained throughout the test sessions. The later models indicated that children with low initial

variability in analogical reasoning profited more from the dynamic-test training than children who were already capable of some analogical reasoning at the first session. Results also showed that spatial working memory was positively related to the number of transformations and the number of explained transformations at pre-test. The MLA results for (explained) transformations and for accuracy differed slightly. For accuracy, results indicated that verbal working memory influenced children's initial status and rate of change. Verbal working memory was, however, not influential for children's initial status or rate of change regarding their (explained) transformations.

Growth-curve trajectories of several subgroups

After running the MLA's, we focused on more in-depth analyses of growth curves of analogical reasoning over time. The results indicated that the most influential variables on children's reasoning performance were: group, spatial working-memory capacity, and initial variability. Spatial working memory was dichotomized into a 'lower score' and 'higher score' category, based on the median score. This allowed us to visually inspect the progression in reasoning performance across sessions for eight subgroups (see Table 7).

Insert Table 7

The differential growth in reasoning performance across sessions of the eight subgroups for the three dependent variables is shown in Figures 3, 4 and 5.

Insert Figures 3,4, & 5

These figures show that children who displayed limited capacity for analogical reasoning at the first session (low variability), made gains after they had received training. Rapid improvement was apparent for all three of the outcome variables. However, these children showed, on average, a dip in performance at the last session for all outcome variables. In contrast, children who displayed greater evidence of analogical reasoning at the first session (high variability) showed a more gradual increase in the quality of their reasoning.

Discussion

The main purpose of the current study was to examine whether dynamic training versus unguided practice resulted in different change paths in strategy-use and in performance accuracy. On the basis of the literature (Halford et al., 2010; Richland et al., 2006) it was anticipated that measures of visual-spatial and verbal working memory and initial variability in children's problem-solving would provide a richer understanding of factors underpinning individual and group growth trajectories in analogical reasoning. Our study also sought to provide a meaningful context in which to investigate the potential value of new ways (both methodological and statistical) to examine inter- and intra-individual change in problem-solving processes within a collaborative testing context. For this reason, our study employed a microgenetic, trial-by-trial assessment to investigate variability in children's strategy-use and to gauge subsequent progress in their use of problem-solving strategies after either intervention or unguided practice.

Our study outcomes revealed that, although repeated practice led to greater accuracy and a superior behavioral strategy-use (with the training adding significantly to this effect), unguided practice alone did not significantly change children's verbal explanations of strategy-use. In contrast, after training, children showed significant improvements in solving the analogy tasks and in both their behavioral strategy-use and verbal accounts. Having tackled each task, each child was asked how they had solved ´the puzzle´. It was found that adaptive (scaffolded)

instruction appeared to be necessary to deepen the children's understanding of how to solve the problems. This deeper understanding seemed to be a necessity before an adequate verbal explanation could be offered (e.g., Crowley & Siegler, 1999) [see also, Rittle-Johnson (2006) for a similar finding]. A related explanation for the limited benefits of unguided practice might be that children in our study had to construct their answers to open-ended tasks. Stevenson, Heiser, & Resing (2016) found that after training, children who had to construct their responses provided higher quality explanations than those who received a multiple-choice format. It seems likely that the combination of training and the construction of the answer helped to deepen the child's understanding.

Our findings only partially confirm those reported by Tunteler et al. (2008) and Tunteler and Resing (2010), who, in their microgenetic studies with classical geometric analogies, found differential progression paths for trained and unguided practice groups, but growth in verbal explanations over time for children in both groups. This may reflect the greater challenge of the items utilized in the present study. Oral explanations of problem solving also reflect metacognitive skills (Efklides, 2006) and do not necessarily parallel children's problem-solving processes, particularly if they are young (Ericsson & Simon, 1984; Veenman, 2015). It seems, therefore, that children can (partially) construct an answer but (still) operate at different levels of verbal and behavioral strategy-use, being able to offer a verbal account only after training (Resing et al., 2012) [see also, Rittle-Johnson, Fyfe, Loehr, and Miller (2015) for related findings in the early numeracy domain].

Contrary to our expectations, initial variability in the child's performance appeared not to be related to change across test sessions (e.g., Siegler, 2007; Tunteler & Resing, 2007). Children with low initial variability in solving analogies profited from the training as well as those who were able to demonstrate a basic level of analogical reasoning at the start of the study. The

training apparently provided them with better tools for constructing and explaining their answers. While dynamic testing appeared to improve the performance of children, irrespective of variability, we were rather surprised to discover that performance dipped at the final measurement, although inspection of individual growth curves also showed large inter- and intra-individual differences in change paths. The growing understanding of how to solve the tasks as evidenced for many of our participants seemed to be time and/or context-bound and did not transfer to a parallel task administered two weeks later. These outcomes may suggest that training was insufficiently extensive, or needed to be provided more frequently over a longer period. Another explanation, supported by the differential patterns of growth-curve trajectories observed, is that specific subgroups of children might require a more tailored form of scaffolded instruction and feedback (Siegler & Svetina, 2002) operating within their zone of proximal development (Alibali & Goldin-Meadow, 1993; Granott, 2002; Wood, Bruner, & Ross, 1976). Of course, some children may have lost interest in having to solve the tasks repeatedly, and may have become less motivated to explain all of the various transformations for every item that they were asked to solve (Siegler & Engle, 1994).

Spatial working memory was as expected found to be related to children's visual-spatial encoding [visual processing being an important component of figural tasks (Laski & Siegler, 2014; Rasmussen & Bisanz, 2005; Van der Ven, Boom, Kroesbergen, & Leseman, 2012)], although this was not associated with improvement on the tasks. In contrast, verbal working-memory positively influenced both children's initial ability and rate of change but only in relation to full construction (accuracy) of the tasks (see also Siegler, 2006). It was not, however, related to the (change) in number of dimensions detected in the task material, or the number and progression of verbal explanations. These findings seem to differ from those of an earlier study by Tunteler and Resing (2010), in which children with a smaller memory span were able to catch

up with their peers after training. While the reasons for this discrepancy are unclear, it should be noted that the tasks in the present study were more demanding in terms of cognitive load and most participants, struggled, especially at the initial stage (Halford et al., 2010). However, the role of working memory (both spatial and verbal) appeared to diminish as discrete aspects of task performance are considered. As children had to construct their answers, it was possible to examine a range of performance variables: the outcome (accurate/inaccurate), the number of objects/features that change and which have to be detected by the child (ranging from 1-6)), and the number of verbal explanations. The graduated prompts training focused on this process of finding changing features in order to solve the analogies. Apparently, as a result of the training, cognitive load diminished, but not sufficiently to completely overcome differences in working memory. An explanation for this finding might be that, of course, a child must start with encoding the task features (visual-spatial), but after this first step, the inductive reasoning process necessarily has more of a verbal character. Training is given orally as well. To be able to analyze, find, and explain the task solution, the child needs sound verbal reasoning (and working memory) ability. In addition, other aspects of higher-order executive functioning not measured in this study, for example, inhibitory control skills (Richland and Burchinal, 2013) appear to play an important role in solving analogies.

Our findings support Siegler's (2006) "overlapping waves" theory, whereby high initial variability of an individual's strategy-use is likely to predict more substantial later learning. Interestingly though, growth trajectories of children with larger (spatial) working-memory capacity, who had received training, suggested that a given level of initial performance was not always necessary for rapid learning to occur. Dynamic testing appeared to be a necessary step for some to learn how to solve the analogy problems: after this, some children displayed a rate of change that exceeded many others.

This study has sought to provide both a methodological and a substantive contribution. Firstly, advanced multilevel analyses were used in combination with a microgenetic design to examine verbal and visual-spatial working-memory components separately (Alloway et al., 2006), and in combination with initial variability and the effects of training versus unguided repeated practice. Secondly, our means of analysis (multilevel analysis for repeated measurement data) enabled us to inspect group-growth trajectories in combination, rather than in isolation, with systematic variation between these trajectories as a function of the background variables and experimental treatment (Van der Leeden, 1998). Furthermore, our analogy tasks were open-ended; children had to construct their answers so enabling us to undertake fine-grained analyses of children's solving processes (e.g., Harpaz-Itay, Kaniel, & Ben-Amram, 2006; Resing & Elliott, 2011; Tzuriel & Galinka, 2000). Although we know that such tasks are more difficult than multiple choice tasks (e.g., Behuniak, Rogers, & Dirir, 1996; In'nami & Kozumi, 2009; Martinez, 1999), and individuals require more help during training to solve these tasks (Stevenson et al., 2016), these type of tasks may improve learning after extensive instruction.

Our study had some limitations. Firstly, our design required a large number of contact hours with the schools and participating children. While many of them were clearly enthusiastic participants, others may have become less motivated or attentive over time, particularly where difficulties were experienced. This was the main reason why children were given just one, rather short training session, which might have led to less than optimal difference between the test outcomes of the trained and non-trained children. Secondly, only two subtests measuring (aspects) of verbal and figural working memory were administered. To strengthen our conclusions regarding the respective roles of verbal and figural working memory in solving analogies before and after training/practice, we would ideally need more measures, based on a memory model built upon empirical relations between working memory, short-term memory, and

general fluid intelligence, for example Engle et al., 1999. A more comprehensive measurement of the verbal memory component may be particularly valuable given the complexity of analogies and inductive reasoning tasks in general. In addition, the current study could only provide some global information regarding the underlying change mechanisms in various subgroups. The relatively small numbers of children in subgroups formed on the basis of working memory and initial variability, for example, did not permit us to offer very fine grained conclusions regarding variability in change over time. As noted above, we clearly saw large individual differences, variable strategy use, and variable learning progression related to working memory, but only when comparing children at a more individual level. Larger training programs (e.g., Tzuriel & George, 2009) with long-term follow-up, and more specific and in-depth measurement of various aspects of working memory are needed to confirm the individual growth curves we found.

Analogical reasoning is of fundamental importance for the understanding, transfer, and retention of many key educational principles in school (Vendetti, Matlen, Richland, & Bunge, 2015). However, laboratory studies suggest that this is not always a spontaneous process (Gick & Holyoak, 1980; 1983) and children can be easily distracted by irrelevant information, for example, perceptual features rather than underlying relationships (Richland et al., 2006). Clearly, the skill of the teacher in supporting the child's use of analogical reasoning is key to effective learning (Richland & Simms, 2015) yet teachers are unlikely to feel confident in how best to support individual children. While we know that differences in the nature and frequency of adult prompts geared to encourage the comparison of analogical relationships can have a significant influence upon children's learning and retention (Vendetti et al., 2015), we still have much to learn about the most effective ways of offering differential instruction that reflects individual differences.

While the current study did not focus in detail on underlying change mechanisms, comparison of the variability and quality of growth-trajectories of different subgroups (such as those identified in the present study) may help to reveal specific strengths and weaknesses that influence particular learning trajectories. This information could be used to better predict children's growth-trajectories and ameliorate potential problems by the use of specialized and targeted instruction or scaffolding. While working memory capacity has been found to influential for all forms of relational reasoning, including that of analogical reasoning, our findings suggested a somewhat complex picture about its relationship to the children's progress. Further research is needed to ascertain which components of working memory are most influential for successful performance and for ongoing development, and whether such factors might be different for high versus low performing children (Grossnickle et al., 2016). Dynamic testing may ultimately contribute to a greater understanding of how differential forms of instruction, ranging from metacognitive to more concrete (Resing, 2000, 2013), can be tailored for children with different cognitive processing profiles.

In a diagnostic context, and in future research, it may prove useful to add a task whereby the child is asked to construct one or more problems for the examiner (in this case a figural analogy) (Bosma & Resing, 2006; Harpaz-Itay et al., 2006; Kohnstamm, 1967). Findings from these studies suggest that such a task may further activate higher-level metacognition, additional strategies and superior explanations.

Within the field of educational psychology, there continues to be significant debate about the value of domain general cognitive assessment for the purposes of informing educational intervention (Compton, Fuchs, Fuchs, Lambert, & Hamlett, 2012; Elliott & Resing, 2015; Fletcher et al., 2011; Fletcher & Vaughn, 2009; Hale et al., 2008, 2010; Reynolds & Shaywitz, 2009). In the eyes of many educationalists and psychologists, psychometric tools and domain-

general approaches have proven valuable for the purposes of prediction and selection, yet continue to offer little to help teachers for making informed decisions about how best to help individual children. It is surely incumbent on educational and cognitive psychologists to devise more sophisticated approaches to understanding individual children's development, and to use this information to inform the design of powerful forms of instruction tailored to individual needs. Analogical reasoning, for example, seems particularly important for school domains such as math (Richland, Zur, & Holyoak, 2007) and reading (Goswami, 2002). Further fine-grained research into the inductive reasoning processes might generate a link to more domain-specific diagnostic research, such as how best to operate response to intervention (RTI) programs. The approach outlined in the present article represents an attempt to make progress in this direction.

Acknowledgements: We would like to thank Mark de Rooij for his helpful comments on an earlier version of this article.

References

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

Alexander, P. A., Willson, V. L., White, C. S., & Fuqua, J. D. (1987). Analogical reasoning in young children. *Journal of Educational Psychology, 79*, 401−408.

Alibali, M., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25*, 468−523.

Alloway, T. P. (2007). *Automated Working Memory Assessment (AWMA)*. London: Harcourt Assessment.

Alloway, T. P., & Gathercole, S. E. (2006). How does working memory work in the classroom? *Educational Research and Reviews, 1*, 134−139.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698−1716.

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ and mathematical skills in children. *Learning and Individual Differences, 21,* 133−137.

Arendasy, M., & Sommer, M. (2005). The effect of different types of perceptual manipulations on the dimensionality of automatically generated figural matrices. *Intelligence, 33,* 307−324.

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

Behuniak, P., Rogers, J. B., & Dirir, M. A. (1996). Item function characteristics and dimensionality for alternative response formats in mathematics. *Applied Measurement in Education, 9*(3), 257−275.

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

Bjorklund, D. F., & Rosenblum, K. E. (2001). Children's use of multiple and variable addition strategies in a game context. *Developmental Science, 4*(2), 184−194.

Bleichrodt, N., Drenth, P. J. D., Zaal, J. N., & Resing, W. C. M. (1984). *Revisie Amsterdamse Kinder Intelligentie Test [Revised Amsterdam Child Intelligence Test]*. Lisse: Swets & Zeitlinger.

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.

Campione, C. J., Brown, A. L., Ferrara, R. S. J., & Steinberg, E. (1985). Breakdown in flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence, 9,* 297−315.

Carr, M., & Schneider, W. (1991). Long-term maintenance of organizational strategies in kindergarten children. *Contemporary Educational Psychology, 16*(1), 61−72.

Chen, Z. (1996). Children's analogical problem solving: The effects of superficial, structural and procedural similarity. *Journal of Experimental Child Psychology, 62,* 410−431.

Chen, Z., & Daehler, M. W. (1989). Positive and negative transfer in analogical problem solving by 6-year-old children. *Cognitive Development, 4,* 327−344.

Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. (2011). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities*, *45*(1), 79−95.

Crowley, K., Siegler, R. S. (1999). Explanation and generalization in young children's strategy learning. *Child Development, 70*, 304−316.

Csapó, B. (1997). The development of inductive reasoning: Cross-sectional assessments in an educational context. *International Journal of Behavioral Development. 20*(4), 609−626.

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, *1*(1), 3−14.

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

Elliott, J. G., & Resing, W. C. M. (2015). Can intelligence testing inform educational intervention for children with reading disability? *Journal of Intelligence, 3*(4), 137−157.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General*, *128*(3), 309-331.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

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*(1), 3−22.

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*, *3*(1), 30−37.

Flynn, E., & Siegler, R. S. (2007). Measuring change: Current trends and future direction in microgenetic research. *Infant and Child Development, 16,* 135−149.

Goswami, U. (2002). Inductive and deductive reasoning. In U. Goswami (Ed.), *Blackwell's handbook of childhood cognitive development* (pp. 282–302). Oxford: Blackwell.

Goswami, U. (2013). The development of reasoning by analogy. In P. Barrouillet & C. Gauffroy (Eds). *The development of thinking and reasoning* (pp. 49-70). Hove: Psychology Press.

Gick, M.L., & Holyoak, K.J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306-355.

Gick, M.L., & Holyoak, K.J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38.Goswami, U. (2002). Inductive and deductive reasoning. In U.

Goswami (Ed.), *Blackwell's handbook of childhood cognitive development* (pp. 282–302). Oxford: Blackwell.

Granott, N. (2002). How microdevelopment creates macrodevelopment: reiterated sequences, backward transitions, and the zone of current development. In N. Grannot & J. Parziale (Eds.), *Microdevelopment: transition processes in developmental and learning* (pp. 80–105). Cambridge: Cambridge University Press.

Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities*, *42*(2), 111−132.

Grossnickle, E.M., Dumas, D., Alexander, P.A., & Baggetta, P. (2016). Individual differences in the process of relational reasoning. *Learning and Instruction*, *42*, 141-159.

Haavisto, M. L., & Lehto, J. E. (2004). Fluid/spatial and crystallized intelligence in relation to domain-specific working memory: A latent-variable approach. *Learning and Individual Differences, 15,* 1−21.

Hale, J. B., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., .. & Simon, J. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly*, *33*(3), 223–236.

Hale, J. B., Fiorello, C. A., Miller, J. A., Wenrich, K., Teodori, A. M., & Henzel, J. (2008). WISC-IV assessment and intervention strategies for children with specific learning difficulties. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (pp. 109−171). New York, NY: Elsevier.

Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Halford, G. S., Wilson, H. W., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental and cognitive psychology. *Behavioral and Brain Sciences, 21,* 803−865.

Halford, G. S., Wilson, H. W., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences, 14*, 497−505.

Harpaz-Itay, Y., Kaniel, S. & Ben-Amram, A. (2006). Analogy construction versus analogy solution, and their influence on transfer. *Learning and Instruction, 16*(6), 582−591.

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

Hessels-Schlatter, C. (2002). A dynamic test to assess learning capacity in people with severe impairments. *American Journal on Mental Retardation, 107,* 340−351.

Holyoak, K. J. (2012). Analogy and relational reasoning. In K.J. Holyoak & R.G. Morrison (Eds.). *The Oxford handbook of thinking and reasoning* (pp. 234-259). New York: Oxford University Press.

Holyoak, K. J., & Nisbett, R. E. (1988). Induction. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of thinking* (pp. 50–91). Cambridge: Cambridge University Press.

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

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

In'nami, Y., & Kozumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing, 26*(2), 219−244.

Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of Educational Research, 78*(1), 85−123.

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

Kolodner, J. L. (1997). Educational implications of analogy: A view from case-based reasoning. *American Psychologist, 52,* 57−66.

Kreft, I., & De Leeuw, J. (1998). *Introducing Multilevel Modeling*. London: Sage Publications.

Kuhn, D. (1995). Microgenetic study of change: What has it told us?.*Psychological Science*, 133-139.

Laski, E. V., & Siegler, R. S. (2014). Learning from number board games: You learn what you encode. *Developmental Psychology, 50,* 853–864.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working-memory in arithmetic problem-solving. *Memory & Cognition, 22,* 395−410.

Martinez, M. E. (1999) Cognition and the question of item format. *Educational Psychologist, 34*(4), 207−218.

Meo, M., Roberts, M., & Marucci, F. S. (2007). Element salience as a predictor of item difficulty for Raven's Progressive Matrices. *Intelligence, 35,* 359−368.

Morrison, R. G., Doumas, L. A. A., & Richland, L. E. (2011). A computational account of children's analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science, 14*, 516−529.

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience, 16*, 260−271.

Perret, P. (2015). Children's inductive reasoning: Developmental and educational perspectives. *Journal of Cognitive Education and Psychology, 14*(3), 389–408.

Primi, R. (2001). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence, 30*, 41−70.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*, 110−122.

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*, 137−157.

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

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., & 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., 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.

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to intervention: Prevention and remediation, perhaps. Diagnosis, no. *Child Development Perspectives, 3*(1), 44−47.

Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science, 24*(1), 87−92.

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.

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *WIRE*s: *Cognitive Science, 6*,177-192.

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science, 316,* 1128−1129.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77,* 1−15.

Rittle-Johnson, B., Fyfe, E. R., Loehr, A. M., & Miller, M. R. (2015). Adding patterns to the equation. *Early Childhood Research Quarterly, 31,* 101−112.

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

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

Siegler, R. S. (2007). Cognitive variability. *Developmental Science, 10*(1), 104–109.

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

Siegler, R. S., & Engle, R. A. (1994). Studying change in developmental and neuropsychological contexts. *Cahiers de Psychologie Cognitive/Current Psychology of* Cognition, *13*(3), 321–349.

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-Freeman, K. E. (2005). Analogical reasoning in 2-year-olds: The development of access and relational inference. *Cognitive Development, 20,* 214−234.

St. Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59,* 745−759.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential.* New York: Cambridge University Press.

Sternberg, R. J., & Grigorenko, E. L. (2006). Culture intelligence and successful intelligence. *Group & Organization Management, 31*, 27−39.

Sternberg, R. J., & Rifkin, B. (1979). The development of analogical reasoning processes. *Journal of Experimental Child Psychology, 27*, 195−232.

Stevenson, C. E., Bergwerff, C. E., 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*(1), 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*(3), 159-169.

Stevenson, C. E., Heiser, W. J., & Resing, W. C. M. (2016). Dynamic Testing of Analogical Reasoning in 5-to 6-Year-Olds Multiple-Choice Versus Constructed-Response Training Items. *Journal of Psychoeducational Assessment*, DOI: 10.1177/0734282915622912

Stevenson, C. E., Resing, W. C. M., & Froma, N. (2009). Analogical reasoning skill acquisition with self-explanation in 7-8 year olds: Does feedback help? *Educational and Child Psychology, 26*, 6−17.

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.

Swanson, H. L. (1994). The role of working memory and dynamic assessment in the classification of children with learning disabilities. *Learning Disabilities Research & Practice*.

Swanson, H. L. (2011). Dynamic testing, working memory, and reading comprehension growth in children with reading disabilities. *Journal of Learning Disabilities, 44*(4), 358-371.

Thibaut, J. P., & French, R. M. (2016). Analogical reasoning, control and executive functions: a developmental investigation with eye-tracking. *Cognitive Development, 38*, 10-26.

Thibaut, J. P., French, R., & Vezneva, M. (2010). The development of analogy making in children: Cognitive load and executive functions. *Journal of experimental child psychology, 106*(1), 1-19.

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 task 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.

Tunteler, E. & Resing, W. C. M. (2010). The effects of self- and other-scaffolding on progression and variation in children's geometric analogy performance: A microgenetic research. *Journal of Cognitive Education and Psychology, 9,* 251−272.

Tzuriel, D., & George, T. (2009). Improvement of analogical reasoning and academic achievements by the *Analogical Reasoning Program* (ARP). *Educational and Child Psychology, 26,* 71–93.

Tzuriel, D., & Flor-Maduel, H. (2010). Theory and research: Prediction of early literacy by analogical thinking modifiability among kindergarten children. *Journal of Cognitive Education and Psychology, 9,* 207−226.

Tzuriel, D. & Galinka, E. (2000). *Children's conceptual and perceptual analogical modifiability test. Instruction manual.* Ramat Gan, Israel: School of Education, Bar-Ilan University.

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

Van der Ven, S. H. G., Boom, J., Kroesbergen, E. H., & Leseman, P. P. M. (2012). Microgenetic patterns of children's multiplication learning: Confirming the overlapping waves model by latent growth modeling. *Journal of Experimental Child Psychology, 113,* 1–19.

Veenman, M. V. (2015). Thinking about metacognition improves thinking. In R. Wegerif, L. Li, & J. C. Kaufman (Eds.), *The Routledge International Handbook of Research on Teaching Thinking* (pp. 280–288). New York: Routledge.

Vendetti, M.S., Matlen, B.J., Richland, L.E., & Bunge, S.A. (2015). Analogical reasoning in the classroom: Insights from cognitive science. *Mind, Brain, and Education, 9*(2), 100-106.

Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (2004). Relational integration, inhibition, and analogical reasoning in older adults. *Psychology and Aging, 19,* 581−591.

Vosniadou, S. (1989). Analogical reasoning as a mechanism in knowledge acquisition: A developmental perspective. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 413−438). Cambridge: Cambridge University Press.

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition, 28,* 1205–1212.

Wickham, H., & Chang, W. (2015). ggplot2: An Implementation of the Grammar of Graphics. R package version 1.0.1, URL http://CRAN.R-project.org/package=ggplot2.

Winne, P. H., & Nesbit, J. C. (2010). The psychology of academic achievement. *Annual Review of Psychology, 61,* 653−678.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89−100.

Appendix A Training procedure (schematic)

Appendix B Regression equations per final MLA model of each analogical performance measure

Step	Instruction	Right Answer	Wrong Answer
0	0.1. Today we are going to make puzzles again. However, this time I will give you some help. 0.2. Just like the other times, there are animals in three boxes, but there are no animals in the fourth box. 0.3. Again please solve this puzzle by putting the animals in this empty box that you think belong there.	1. Yes, that's correct. 2. How did you solve the puzzle? / Why did you put these animals here? ***Go to step 6***	1. Your solution isn't completely correct yet. 2. I will put the cards back and give you some help.
1	1.1. First, you think about where to start. *1.2. Experimenter tells child which boxes of the puzzle belong together.* 1.3. What do you think should be put in the empty box?	1. Yes, that's correct. 2. How did you solve the puzzle? / Why did you put these animals here? ***Go to step 6***	1. It's not completely correct yet. 2. I will give some more help. [*experimenter puts the cards back*]
2	2.1. You can solve the puzzle by following these steps. 2.2. First, you compare the boxes and then you think how the boxes belong together. 2.3. Then you put the animals in this empty box that you think belong there. 2.4. When you have solved the puzzle, you can check if your answer is correct by comparing the boxes.	1. Yes, that's correct. 2. How did you solve the puzzle? / Why did you put these animals here? ***Go to step 6***	1. It's (almost/ not completely) correct yet. 2. Let's look at it together. [*experimenter puts the cards back*]
3	3.1. We start by comparing the boxes. *3.2 Experimenter asks child to explain changes from A to B and from A to C.* *3.3. Experimenter points out the similarity between [C:D] and [A:B], and between [B:D] and [A:C].* 3.4. So, how do we fill the empty box to solve the puzzle? *3.5. Experimenter asks child to check whether his/her answer is correct.*	1. Yes, that's correct. 2. How did you solve the puzzle? / Why did you put these animals here? ***Go to step 6***	1. Ok, (it's almost correct) I shall give you some more help. [*experimenter puts the cards back*]
4	4.1. This box [A] changed to that box [B] because.... *[Experimenter explains all transformations from A to B]* 4.2. This box [B] changed to that box [D] because.... *[Experimenter explains all transformations from B to D*]	1. Yes, that's correct. 2. How did you solve the puzzle? / Why did you put these animals here? ***Go to step 6***	1. We are going to solve the puzzle together. [*experimenter puts the cards back*]
5	5.1. We start with the animals. Which animals do we need? *If answered incorrectly: Experimenter gives correct answer and explains animal type changes in [A:B] and [B:D].* 5.2. Which color [animal(s)] do we need / and which color [animal(s)]? *If answered incorrectly: Experimenter gives correct answer and explains color changes in [A:B] and [B:D].* 5.3. Do we need large or small [animal(s)] / and large or small [animal(s)]? *If answered incorrectly*: *Experimenter gives correct answer and explains size changes in [A:B] and [B:D].* 5.4. Do we need one or two [animal(s)] / and one or two [animal(s)]? *If answered incorrectly*: *Experimenter gives correct answer and explains quantity changes in [A:B] and [B:D].* 5.5. In which direction do the [animal(s)] / and [animal(s)] need to walk? *If answered incorrectly*: *Experimenter gives correct answer and explains direction changes in [A:B] and [B:D].* 5.6. Do we place the [animal(s)] at the top or the bottom of the empty box / and the [animal(s)]? *If answered incorrectly*: *Experimenter gives correct answer and explains position changes in [A:B] and [B:D].*	[*In partnership with the child the right answer is created. Each sub-step needs to be mentioned]* 1. That is correct! 2. And why is this correct? [*experimenter continues to request information until the child gives no more information*]	
6	*Give the correct explanation for the answer*.		

Regression Equation
Solutions Correct = 1.09 + 8.13 variability + 0.03 spatial working memory + 6.31 condition + 0.03 verbal working memory*session– 1.88 condition*variability – 1.47 session*condition.
Transformations Correct (in behavioral solutions) = 24.72 + 2.28 session + 47.99 variability + 32.75 condition – 15.33 condition*variability – 4.45 session*condition.
Transformations Explained = 3.99 + 29.84 variability + 0.11 spatial working memory + 23.82 condition – 6.97 condition*variability – 5.54 session*condition.

Note. All variables contain a meaningful 0-point (including session).

Figure 1a

Figure 1

Examples of figural analogies used during non-guided practice and dynamic testing sessions (adopted from Stevenson et al., 2009)

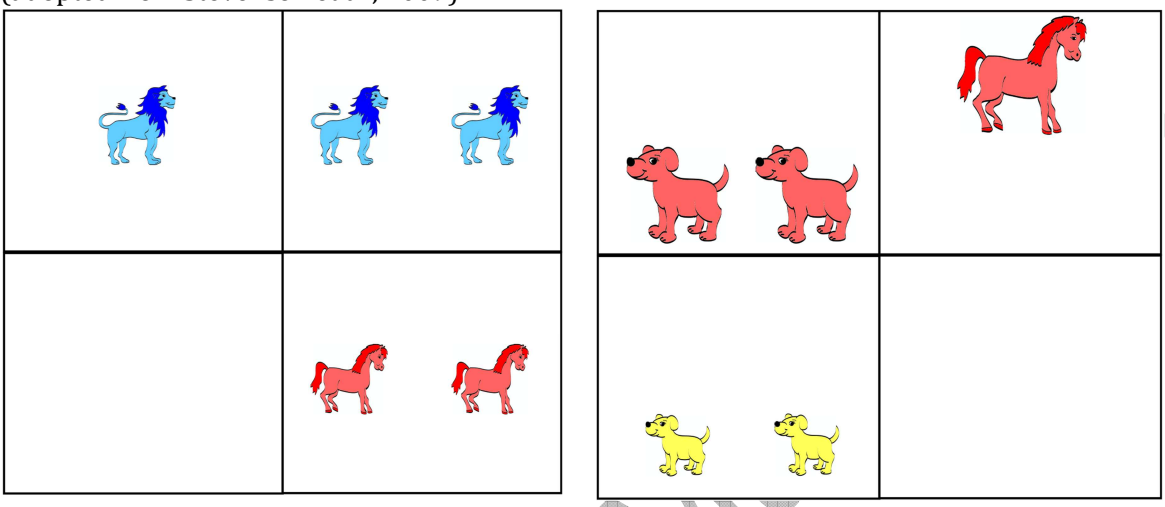

Figure 2

Progression in the number of correct solutions (upper figure); change patterns (middle figure) in the number of correctly solved transformations; and the number of transformations explained across sessions (lower figure), for the training and repeated practice groups respectively.

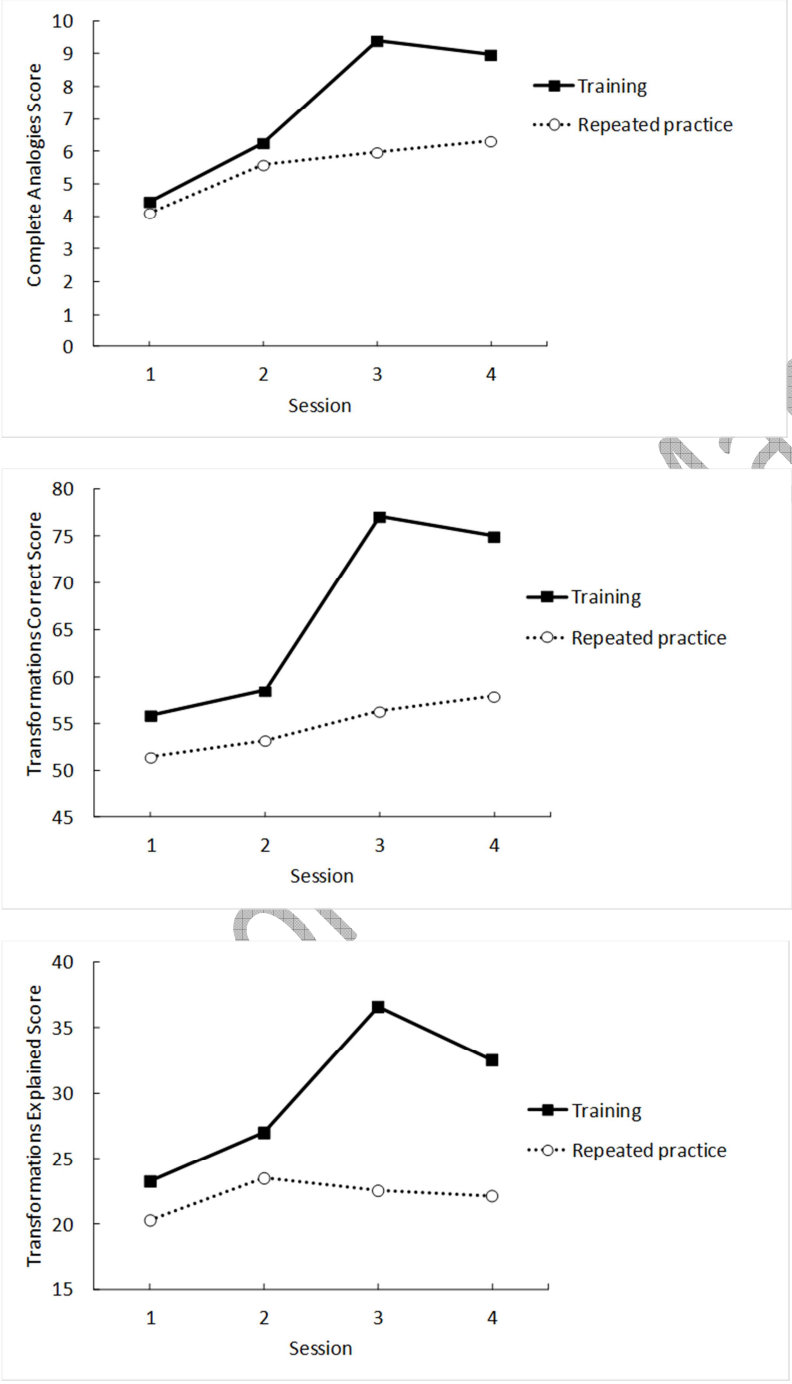

Figure 3

Number of complete solutions for the eight subgroups across sessions

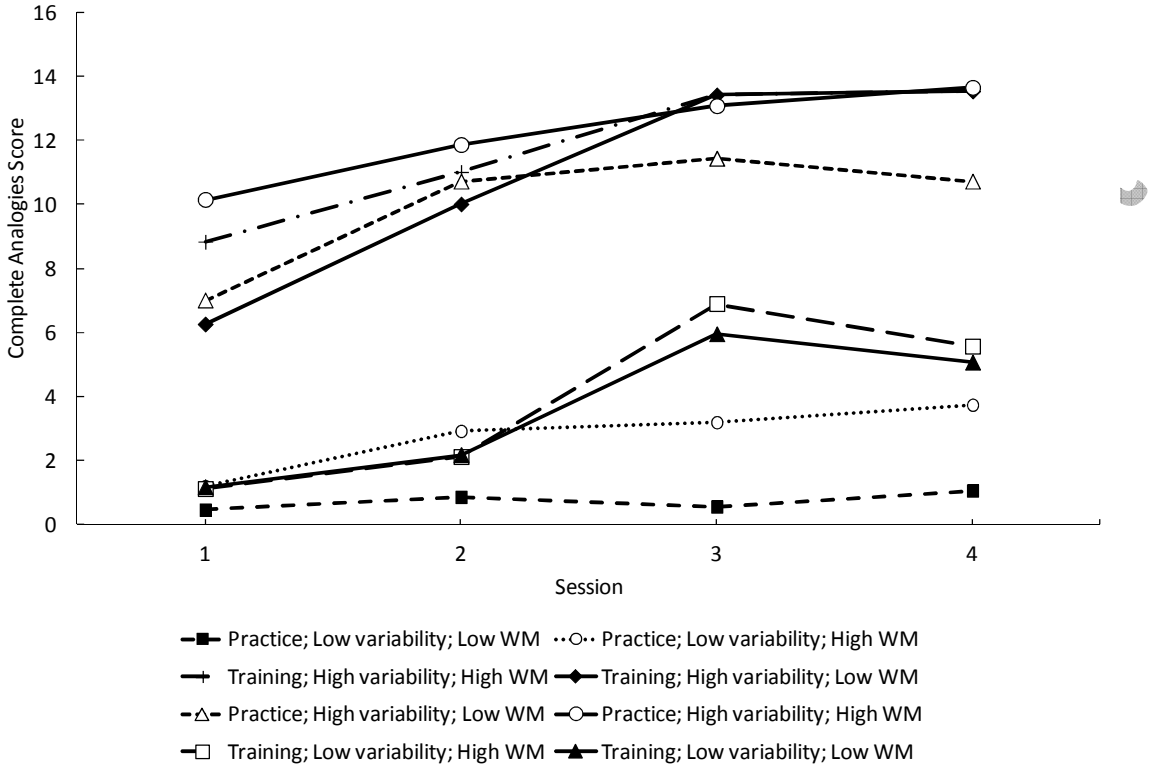

Figure 4

Number of correct transformations for the eight subgroups across sessions

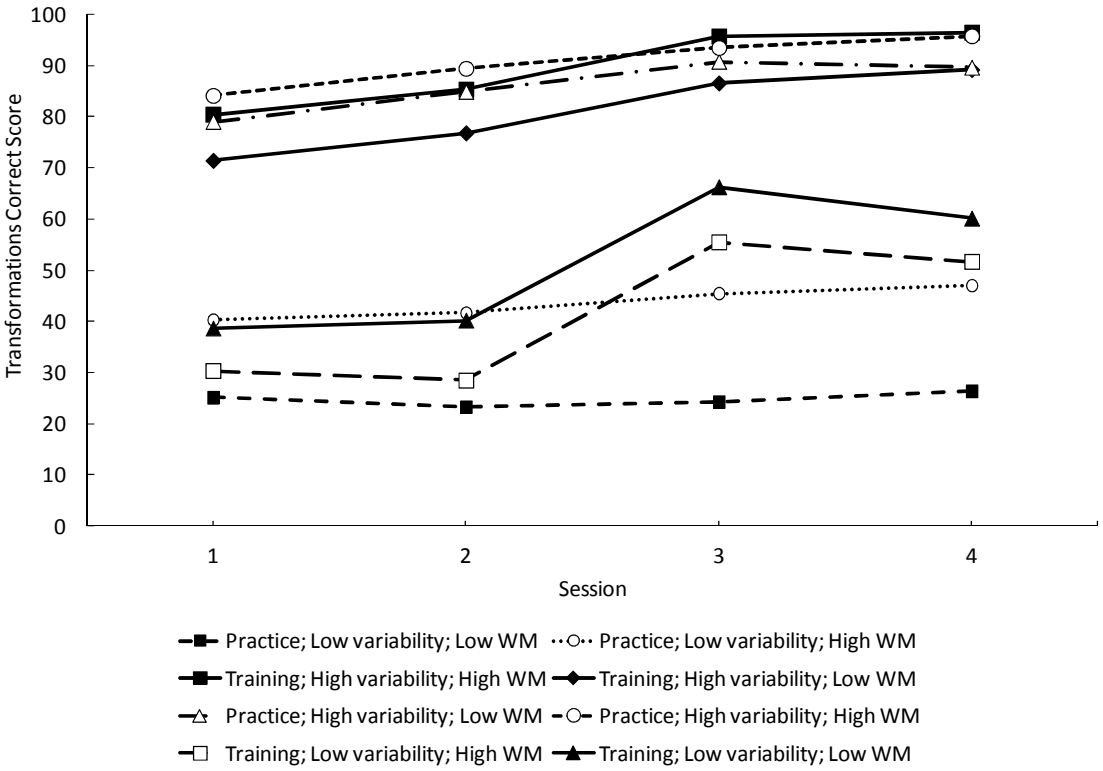

Figure 5

Number of explained transformations for the eight subgroups across sessions

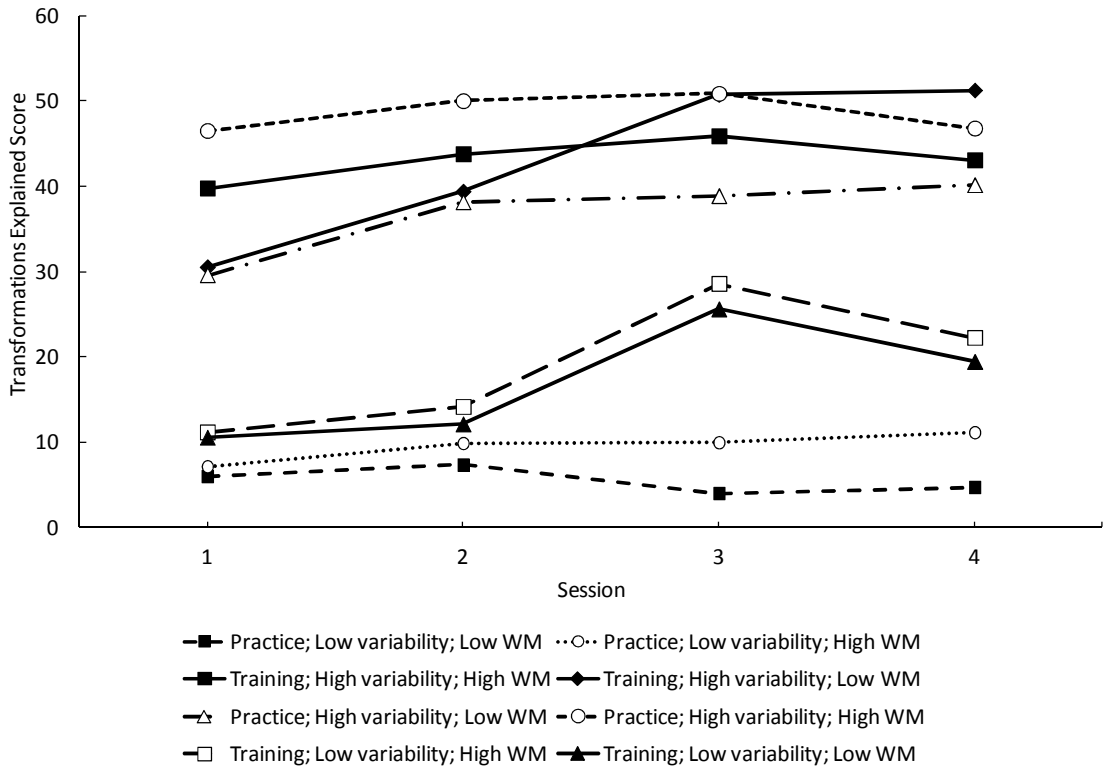

Table 1. Design of the study. Dynamic training (DT) Session: the practice-condition received the same items as the training condition (x), but without training (T)

Condition	Session					
	Pretesting	1	2	DT	3	4
Practice	X	x	x	x	x	x
Training	X	x	x	T	x	x

Table 2. Means and standard deviations of analogy scores per session and condition

	Session			
	1	2	3	4
Condition	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
	Complete Solutions (Score range 0-20)			
Practice (N = 52)	4.10 (4.83)	5.6 (5.56)	5.94 (6.28)	6.31 (6.1)
DT (N = 52)	4.44 (4.33)	6.25 (5.66)	9.38 (5.35)	8.96 (5.61)
Total (N = 104)	4.27 (4.57)	5.92 (5.59)	7.66 (6.06)	7.63 (5.98)
	Correct Transformations (Score range 0-110)			
Practice (N = 52)	51.44 (31.99)	53.88 (35.44)	56.71 (37.09)	58.35 (36.8)
DT (N = 52)	55.88 (28.32)	59.65 (31.97)	78.37 (25.6)	76.08 (26.73)
Total (N = 104)	53.66 (29.63)	56.77 (33.71)	67.54 (33.53)	67.21 (33.28)
	Explained Transformations (Score range 0-110)			
Practice (N = 52)	20.29 (21.11)	23.48 (24.2)	22.54 (24.31)	22.13 (22.69)
DT (N = 52)	23.25 (17.39)	26.98 (21.38)	36.6 (22.25)	32.52 (23.51)
Total (N = 104)	21.77 (19.31)	25.23 (22.79)	29.57 (24.24)	27.33 (23.56)

Table 3. Means and standard deviations of verbal and spatial working memory per condition

	AWMA Verbal	**AWMA Spatial**
Condition	Mean (SD)	Mean (SD)
Practice (N = 52)	28.29 (10.61)	47.38 (22.90)
DT (N = 52)	27.38 (8.96)	49.69 (24.20)
Total (N = 104)	27.84 (9.78)	23.47 (23.47)

Table 7. Subgroups of children based on the variables condition, spatial working memory and initial variability

Group Code[1]	**000**	**001**	**010**	**011**	**111**	**110**	**101**	**100**
Condition	0	0	0	0	1	1	1	1
Spatial working memory	0	0	1	1	1	1	0	0
Initial variability	0	1	0	1	1	0	1	0
Number of children	20	7	11	14	17	9	8	18

Note. [1]Group codes are based on condition: 0 = repeated practice, 1 = training; spatial working memory: 0 = lower, 1 = higher; initial variability: 0 = low, 1 = high.

Table 4. Results of multilevel modeling with predictors of the intercept and slope of the number of complete analogies

	M0	M1	M2	M3	M4	M5	M6	M7	M8
Fixed Effects									
Intercept	6.370***	4.593***	4.573***	-1.213	-0.574	-0.776	1.063	1.078	1.086
Time		1.185***	0.902***	0.897***	-0.368***	0.131	0.051	0.036	-0.155
Train			5.412***	5.428***	5.417***	4.652**	5.407***	5.413***	6.312***
Time x Train			-1.454***	-1.452***	-1.429***	-1.452	-1.449***	-1.460***	-1.470***
WMV				0.050	0.034	0.034	-0.060*	-0.056*	-0.064*
WMS				0.091***	0.086***	0.090***	0.034**	0.034**	0.035**
Time x WMV					0.028*	0.028.	0.031*	0.026**	0.038**
Time x WMS					0.009	-	-	-	-
Condition x WMV						0.008	-	-	-
Condition x WMS						0.011	-	-	-
Initial performance							7.933***	7.707***	8.125***
Time x Initial performance								0.355	-
Condition x Initial performance									-1.858**
Variance Components									
Within-person (Residual)	7.98	3.82	3.28	3.28	3.28	3.28	3.28	3.28	3.15
In initial status	24.82	20.32	20.69	16.42	16.39	16.38	4.21	4.20	4.38
In rate of change		1.10	1.01	1.01	0.87	0.90	0.92	0.90	0.98
Goodness-of-fit									
Deviance	2315	2161	2118	2094	2085	2088	1968	1967	1961
AIC	2320	2173	2134	2114	2109	2114	1992	1993	1987
BIC	2333	2197	2166	2154	2157	2166	2040	2045	2040
Number of Parameters	3	6	8	10	12	13	12	13	13

Note. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Deviance, AIC, and BIC statistics were assessed to compare the relative goodness-of-fit of the successive models.

Table 5. Results of multilevel modeling with predictors of the intercept and slope of the number of correct transformations

	M0	M1	M2	M3	M4	M5	M6	M7	M8
Fixed Effects									
Intercept	60.670***	53.296***	53.680***	21.511*	26.264*	22.189*	25.956***	25.855***	24.721***
Time		4.916***	2.283*	2.239*	-1.812	2.238*	2.286*	2.388*	2.287*
Train			25.400***	25.402***	25.409***	22.333**	25.400***	25.384***	32.745***
Time x Train			-4.450*	-4.369*	-4.383*	-4.467*	-4.451	-4.439*	-4.45*
WMV				0.267	0.154	0.238	-	-	-
WMS				0.510[a]	0.477***	0.512*	0.156[c]	0.156.	0.159.
Time x WMV					0.096	-	-	-	-
Time x WMS					0.028	-	-	-	-
Condition x WMV						0.137	-	-	-
Condition x WMS						-0.017	-	-	-
Initial performance							45.527***	45.757***	47.987***
Time x Initial performance								-0.241	-
Condition x Initial performance									-15.334***
Variance Components									
Within-person (Residual)	207.90	99.72	81.55	81.56	81.56	81.73	81.55	81.55	76.11
In initial status	909.50	893.11	907.45	780.08	778.10	779.48	322.01	322.00	330.26
In rate of change		40.71	37.22	37.21	35.78	36.70	37.22	37.20	38.27
Goodness-of-fit									
Deviance	3704	3587	3531	3512	3509	3511	3425	3424.7	3408
AIC	3710	3599.0	3546.8	3532	3533	3535	3445	3446.7	3430
BIC	3722	3623.2	3579.0	3572	3581	3583	3485	3491.0	3474
Number of Parameters	3	6	8	10	12	12	10	11	11

Note. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Deviance, AIC, and BIC statistics were assessed to compare the relative goodness-of-fit of the successive models.

Table 6. Results of multilevel modeling with predictors of the intercept and slope of the number of explained transformations

	M0	M1	M2	M3	M4	M5	M6	M7	M8
Fixed Effects									
Intercept	25.974***	22.822***	22.742***	-1.596	-1.544	-1.631	4.079	4.103	3.985
Time		2.101***	1.043	1.050	0.894	1.040	1.044	0.986	1.054
Train			20.456***	20.512***	20.535***	22.029***	20.452***	20.456***	23.815***
Time x Train			-5.515***	-5.548***	-5.594***	-5.526***	-5.514***	-5.521***	-5.538***
WMV				0.286	0.273	0.285	-	-	-
WMS				0.338***	0.344***	0.285***	0.116*	0.116*	0.114*
Time x WMV					0.038	-	-	-	-
Time x WMS					-0.018	-	-	-	-
Condition x WMV						0.090	-	-	-
Condition x WMS						-0.086	-	-	-
Initial performance							29.474***	29.416***	29.838***
Time x Initial performance								0.138	-
Condition x Initial performance									-6.968*
Variance Components									
Within-person (Residual)	95.11	60.21	52.22	52.22	52.19	52,23	52.22	52.22	50.60
In initial status	417.16	363.13	368.50	293.11	293.07	293.39	117.58	117.58	119.62
In rate of change		16.53	15.77	15.78	15.52	15.25	15.77	15.77	16.31
Goodness-of-fit									
Deviance	3379	3317	3278	3257	3258	3254	3175	3175	3169
AIC	3385	3329	3294	3277	3280	3279	3195	3197	3191
BIC	3397	3353	3326	3317	3324	3327	3236	3242	3236
Number of Parameters	3	6	8	10	12	12	10	11	11

Note. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Deviance, AIC, and BIC statistics were assessed to compare the relative goodness-of-fit of the successive models.