Dynamic testing and transfer:

An examination of children's problem-solving strategies

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

Word count: (text [excl. acknowledgement])

Correspondence concerning this paper should be addressed to Prof. dr. Wilma C. M. Resing, Faculty of Social Sciences, Department of Psychology, Section Developmental and Educational Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands. Tel. (+31) 71.527.3680/ or 3644.

Email address: resing@fsw.leidenuniv.nl

Abstract (150 words max)

This study examined the problem-solving behaviour of 104 children (aged 7-8 years) when tackling construction-analogy tasks. Children were allocated to one of two conditions: either a form of unguided practice alone or this in combination with training based on graduated prompt techniques. Children's ability to solve figural open-ended analogy-problems was investigated as well as their ability to construct new analogy problems themselves. We examined children's progression in solving analogy problems and the variability in their strategy-use. Results showed that the group that received training made greater progress in solving analogy problems than children who only received unguided practice opportunities. However, the training appeared to give no additional improvement in performance on the transfer task over that of repeated unguided practice alone. Findings from this study demonstrate that an open construction task can provide additional information about children's cognitive learning potential.

Dynamic testing and transfer:

An examination of children's problem-solving strategies

Resing, W.C.M., Bakker, M., Pronk, C.M.E., & Elliott, J.G.

Introduction

In education, teaching new concepts, knowledge, and problem-solving strategies is common practice. A core aim of teaching is transfer of learning to new situations and contexts. The ability to generalize that what has been learned to new, related tasks, however has been and still is a major challenge (Bransford & Schwarz, 1999; Day & Goldstone, 2012; De Corte, 2003), and has been subject of research for more than a century (Engle, 2012; Larsen-Freeman, 2013). According to Holyoak (1984), transfer requires individuals to perceive the underlying relationship between two problems with some level of similarity (see also Brown, 1982; Brown, Bransford, Ferrara, & Campione, 1983). This is often difficult to accomplish, particularly for young children, and many cognitive training studies since then have shown that children do not easily generalize newly learned strategies to other problems and contexts (e.g., Detterman, 1993; Opfer & Thompson, 2008). Holyoak described transfer as the process of finding an analogy between a base (trained task) and a target problem (transfer task). This process will end unsuccessfully "if the problem solver fails to encode elements of the schema, in either the base or the target"[problem] (Holyoak, 1984, p. 218), or if a taught strategy becomes "welded" to a specific task (Brown, 1978). Furthermore, the use of taught knowledge or problem-solving strategies is often restricted to near transfer and limited to (very) familiar contexts and purposes (e.g., Detterman, 1993).

Siegler (2006) noted that children will only show transfer of knowledge to novel tasks once they have become good strategic solvers, although this is, according to him, likely to be preceded by variable strategic behaviour (see also Perry, Samuelson, Malloy, & Schiffer, 2010). With reference to dimensions such as content and context (Barnett & Ceci, 2002), researchers have differentiated, among others, between surface versus deep transfer (Forbus, Gentner, & Law, 1995), formal versus material transfer (Klauer, 1998), and near versus far transfer (Jacobs & Vandeventer, 1971). Transfer has been found to occur consciously and unconsciously (Day & Gentner, 2007; Day & Goldstone, 2012), instantaneously and very gradually, and after task mastery (Siegler, 2006), or after more variable strategic behaviour (Perry et al., 2010).

Variability in performance and strategy-use on one or more tasks occurs both over the course of development and across cognitive domains. This has been demonstrated by several studies regarding the range of cognitive strategic behaviours that children show when solving scholastic tasks (e.g., Bjorklund & Rosenblum, 2001; Chen & Siegler, 2000; Siegler, 2007; Tunteler & Resing, 2002, 2010). Therefore, when measuring children's cognitive abilities and potential to generalize learned knowledge and solving procedures to new, related tasks, it would be interesting to take into account children's progression as well as fluctuations in their use of problem-solving strategies when tackling tasks.

Inductive reasoning tasks, such as classification, analogies, or series completion, are often used in the measurement of cognitive development and transfer. The ability to reason by analogy, for instance, is even thought to be closely related or even identical to the underlying transfer process (Alexander & Murphy, 1999; Reeves & Weisberg, 1993). Inductive reasoning tasks all require comparable underlying problem-solving processes: starting with specific observations of the provided information, a rule that leads to the solution must be detected and formulated. This rule finding process is achieved through comparison processes (Holyoak & Nisbett, 1988; Pellegrino & Glaser, 1982; Perret, 2015). Klauer (1992, 2014) stated that all inductive tasks can be solved by means of systematic comparison processes, which involve finding similarities and/or differences between task attributes and/or relations among attributes. Inductive reasoning undoubtedly plays an important role in (classroom) learning processes as well as in transfer, as these often also require the ability to detect regularities in seemingly non-ordered material and rules for task solving (e.g., Csapó, 1997; Csapó, Molnár, & Nagy, 2014; Goswami, 1992; Klauer & Phye, 2008; Morrison et al., 2004; Perret, 2015; Vosniadou, 1989). Close inspection of children's inductive reasoning abilities, including the influence of training on the use of more advanced solving strategies (e.g., Higgins, 2015), is complex as a consequence of the variability in performance over time, as sketched above. Transfer of analogical reasoning skills to new tasks or contexts is also difficult to detect, because children show individual differences in the development of their strategy-use over time (Siegler, 2006).

The current study sought to examine children's strategic behaviour when tackling different types of construction-analogy tasks: those where tasks were presented to them and others where they were asked to construct new problems themselves. Children received either a form of unguided practice alone or this in combination with training. By these means, we compared the effects of two different treatments on children's solving of tasks requiring analogical reasoning and on their ability to generalize knowledge and procedures derived from (one of) these treatments to new tasks. The main focus of our study was therefore on 'the breadth of change' dimension of Siegler's (1996)

'overlapping waves' theory. This theory refers to the range of change, variability, generalization or transfer of previous learning to other problems and contexts.

Our study differed from most other studies on transfer because a 'reversal' procedure was employed. As mentioned, in addition to opportunities to practice, some of the children received a training session geared towards helping them to solve open-ended figural analogy problems. Subsequently, the children were also invited to take a more active role by constructing new problems of the kind they had been given before, which the examiner was then required to solve (Bosma & Resing, 2006; Kohnstamm, 2014).

To encourage transfer of previously learned or practiced problem-solving strategies, the surface features of our reversal (self)construction task were kept the same as those of the open-ended figural analogies tasks children had to tackle during the practice and training sessions. During these sessions, children had to construct and subsequently explain their answers. We assumed that the use of the same matrix-format and testing materials would prime the children to draw upon their previous experiences and learning (Day & Goldstone, 2012). This permitted the construction of comparable/equivalent figural analogies as were administered in the tasks used during the previous sessions. Nevertheless, these surface similarities do not necessarily make the process of transfer straightforward. The reversal (self)construction format was assumed to be much more challenging than the open-ended figural analogies task, since it was assumed that the former required children to extract analogical strategies from their own idiosyncratic schemas in their memory in order to construct the analogies. Such complexity would not be required for the tasks with the "normal" open-ended construction format (Martinez, 1999*)*. Our transfer task was assumed to balance between the "high road" and "low road" transfer mechanisms discerned by Salomon and Perkins (1989). According to them, high-road transfer occurs through intentional mindful abstraction of elements from a certain context including their application to a new context. The elements abstracted during this process often have the form of a rule or scheme. Low-road-transfer can only lead to a narrow range of transfer and is based on extensive and varied practice and occurs as a consequence of automatic activation of previously learned behaviour in a new situation.

Providing children with the opportunity to move beyond practice experiences to engagement in problem construction may not only shed light on children's abilities to transfer learning but also on individual differences in the developing use of problem-solving strategies (Haglund & Jeppsson, 2012; Kim, Bae, Nho, & Lee, 2011; Pittman, 1999; Siegler, 2006). Therefore, the problem construction task utilised in the current study was administered to assess the extent to which children's learning on the initial task subsequently transferred to one that involved a reversal of roles. However, to achieve this, an in-depth understanding of children's growth trajectories would be useful. Here, the use of a

microgenetic research design may prove especially helpful (Siegler, 2006; Siegler & Svetina, 2002; Tunteler & Resing, 2007; Winne & Nesbit, 2010). Such measurement designs, including regular measurements of an individual's changing performance on a number of parallel tasks over a relatively short period of time (Siegler & Crowley, 1991), were developed to investigate both the spontaneous development of cognitive abilities and a child's learning progress. In the current study, we combined both dynamic testing and a microgenetic method of studying children's development and learning. We investigated whether children's growth trajectories showed differing pathways of transfer when acquired through more 'natural' unguided practice opportunities than when a short training procedure, as part of a dynamic test, was included.

Dynamic testing, using a test-training-test format, has become increasingly used for the study of inductive reasoning (e.g., Bethge, Carlson, & Wiedl, 1982; Budoff, 1987; Ferrara, Brown, & Campione, 1986; Resing, 1993; Resing & Elliott, 2011; Tzuriel, 2013). Key to this approach is the incorporation of feedback and training during the testing phases (Elliott, Grigorenko, & Resing, 2010; Grigorenko & Sternberg, 1998: Haywood & Lidz, 2007). Unlike conventional forms of static testing, where usually no feedback on how to improve performance is given, dynamic testing aims to investigate children's progression in performance after they have been given explicit, sometimes tailored, assistance during test session(s). Such a testing procedure may provide important information about children's potential for learning (Grigorenko, 2009; Jeltova et al., 2011; Resing, 2013). In fact, a (dynamic) training procedure combined with a microgenetic research method has been found to yield significant differential inter- and intra-individual learning trajectories after both repeated practice, and training experiences (Resing, 2013; Resing, Tunteler, & Elliott, 2015; Tunteler, Pronk, & Resing, 2008). In the current study, we used a training procedure derived from dynamic testing to examine whether training and practice, versus practice alone, resulted in different growth trajectories in children's analogical reasoning and ability to transfer the learned or practiced solving strategies to a new, but strongly related, (self) construction task.

To aid our analysis, we included children's explanations immediately gathered after administering each analogy problem (Church, 1999; Siegler & Stern, 1998). For children aged five years and older, the literature has shown the benefit of combining observations of behavioural solution strategies with their immediate descriptions of how they sought to solve the problem. The value of this approach has been shown in studies in various domains, such as arithmetic (Siegler & Stern, 1998), reading (Farrington-Flint, Coyne, Stiller, & Heath, 2008), and inductive reasoning (Resing, Xenidou-Dervou, Steijn, & Elliott, 2012; Stevenson, Hickendorff, Resing, Heiser, & de Boeck, 2013).

In the current study, an innovative way of measuring transfer was used, derived from Kohnstamm's (1968, 2014) procedure for testing young children's ability to solve seriation/class inclusion tasks. The study examined whether children were able to construct transfer tasks themselves after repeated unguided practice. Comparison on this was undertaken between two groups, one of which had received relatively brief training on how to solve figural analogy tasks. We anticipated that a microgenetic study design would provide us with a more forensic account of children's progress in this domain. We expected children who progressed most in solving the analogies after the practice or training session, would be more able to construct new analogy tasks themselves, or at least tasks that reflected a rudimentary notion of analogical reasoning.

A number of hypotheses were tested. Although children's performance in relation to the construction of transfer tasks was the primary focus of our research, we were also interested in examining if their responses improved as a consequence of repeated practice. Firstly, we examined whether young children, when given opportunities for repeated practice including or not including training, adopted more advanced solving strategies, measured in terms of their a) accuracy, b) use of task-solving components, and c) verbal explanations (e.g., Tunteler & Resing, 2010; Resing et al., 2015). Based on prior research, we hypothesised that performance on these measures would change as a consequence of practice alone and that children in the training condition would outperform children in the unguided practice condition on all measures; we also expected large individual differences (e.g., Resing & Elliott, 2011).

A second set of hypotheses concerned the number of accurately constructed reversal analogies, the percentage of accurate task-solving components and children's explanations during the transfer tasks. We expected that these three measures would be related to condition (repeated practice and training or repeated practice alone), progress in both accuracy in solving analogies and quality of the strategies employed. We expected that children would only show transfer once they had become good strategic problem solvers (Siegler, 2006). Optimal performance on inductive reasoning tasks is strongly influenced by strategy-use (Siegler & Svetina, 2002), and training has been shown to increase children's ability to explain the changes that occur within the tasks (Resing et al., 2012). Therefore, those children who were able to explain their problem-solving behaviour at the end of the study period were expected to construct the reversal transfer tasks in a particularly effective manner They are likely to have gained a good understanding of the underlying principles of the figural analogies (Harpaz-Itay, Kaniel, & Ben-Amram, 2006; Perkins, 1992), which should enable them to apply this knowledge to the transfer task.

Finally, we explored inter-individual differences in children's accounts of how they had tried to tackle the transfer task. We expected that they would make most reference to changes in colour,

size, and number (e.g., Mulholland, Pellegrino, & Glaser, 1980). Children and refer less frequently to the more complex task-solving components - orientation and positions. However, we expected that children who had received training would mention all the components, more frequently than those in the unguided practice condition (Stevenson, Bergwerff, Heiser, & Resing, 2014). We also took into account children's overall quality and progression of their solution strategies during the unguided practice sessions. We expected that children demonstrating higher strategic ability would make reference to a greater number of task components.

Method

Participants

The study included 104 children attending the second grade of primary school. Participants were 53 girls and 51 boys, aged 7-8 years with a mean age of 7.8 years (*SD* = 4.8 months; age range 7.0-8.7 years). The children were recruited from eight schools in the Netherlands that were selected on the basis of their willingness to participate in the study. The children represented a full range of socioeconomic backgrounds. Dutch was the first language spoken at school and at home. Data collection was undertaken by postgraduate students with experience of educational and dynamic testing. Parental consent for participation was obtained in all cases. All children participated in each test session.

Design

The study employed a two-pre-test-two-post-test-two-experimental-groups randomized block design, with a blocking procedure based on Exclusion, a subtest measuring visual inductive reasoning derived from a Dutch child intelligence test, the RAKIT (Bleichrodt, Drenth, Zaal, & Resing, 1984). Before the administration of the first pre-test, children were, on the basis of this blocking procedure, randomly allocated to one of two treatment conditions: a (static) 'unguided practice' condition or a (dynamic) 'unguided practice plus training' condition (see the design scheme in Table 1). Unlike children in the unguided practice group, children in the training condition received training between the two unguided pre-test and post-test practice sessions. In that period, children in the unguided practice condition received practice-items without any extra instruction. In all other respects they received the same inputs as the children of the training condition. All unguided practice sessions took around 30 minutes per child, whereas the training session took 30-60 minutes per child. After the final unguided practice session, children of both treatment groups received the same reversal construction task, which was used as a measure of transfer.

- Insert Table 1 –

Instruments

Exclusion. Exclusion is a visual inductive reasoning subtest of a Dutch child intelligence test: RAKIT (Bleichrodt et al., 1984). The test consists of 40 items each comprising four geometric figures. Three of the figures can be grouped together according to a rule that needs to be identified. The test requires the child to select the figure that, in each case, does not fulfil the rule. The test is considered to tap children's ability to infer rules, an important prerequisite for successful inductive reasoning.

Figural analogies. The figural analogical reasoning tasks used in this study consisted of an adapted, open-ended response version of the concrete figural dynamic test Animalogica, developed by Stevenson and Resing (Stevenson, Heiser, & Resing, 2016; Stevenson, Touw, & Resing, 2011). To solve the figural matrix analogies (A:B = C: ? arranged in a 2x2 matrix, see Figure 1), children had to encode item attributes and infer a relationship between two given animal pictures of an analogy (A, B) and then apply this to the third picture (C) of the analogy. Finally, a fourth animal picture had to be found (from a range of possibilities), so that the relationship between the third and fourth picture equalled that of the first two pictures (e.g., Sternberg, 1985). To construct their response, children were asked to choose the accurate card from 72 coloured cards, entailing six types of animals with three familiar colours, and two sizes. The cards were printed on both sides with two copies of each image, enabling the child to transform both quantity and orientation of the animals when solving the analogy. Four parallel item sets with 20 open-ended tasks each were systematically designed to appear different by changes in animal type and the colour of the figures but at the same time having equal item difficulties.

- Insert Figure 1 -

Figural analogies training. The short dynamic testing procedure used in this study is known as a graduated prompts approach. The training procedure was originally pioneered in a dynamic-testing setting by Campione and Brown (e.g., Campione, Brown, Ferrara, Jones, & Steinberg, 1985), and has been successfully further developed and utilized in studies on dynamic testing (e.g., Fabio, 2005; Resing, 2000; Resing & Elliott, 2011). The procedure involved the use, during the training session, of a series of adaptive and standardized, hierarchically ordered, prompts that proceed from general, metacognitive, to increasingly task-specific prompts. The prompts were only provided when a child was unable to proceed independently. Thus, the children were provided with the minimum

number of prompts possible to enable progression through the test. The items for training consisted of a set of 7 concrete figural analogies similar to those employed in the other sessions (adapted from Resing, 2000). Table 2 shows a schematic overview of the prompts procedure.

- Insert Table 2 -

Construction tasks (and role-reversal) procedure. Children had to construct three analogy tasks themselves and then had to reverse roles and administer these tasks to the examiner. The first construction task included both an A4-sized sheet displaying an (2x2) analogies matrix with four empty quadrants and some baskets with all 72 animal cards (small, large, various colours; two-sided). The children could therefore construct a variety of different analogies. They were informed that the roles were going to change: they would now be the teacher and the examiner would take on the role of the child. The child was shown the empty matrix for the first construction task (I) and was told that what they saw was an 'empty puzzle' in which he or she was allowed to make a puzzle using any of the cards, just like the puzzles they had solved before. Now, however, they were the teacher and the examiner was the child. We thought this procedure would give children the opportunity to spontaneously display their understanding of analogies and analogical/inductive reasoning so that at least some of them would be able to display transfer of their newly learned strategies. After receiving the 'puzzle' made by the child, the examiner filled the empty quadrant of the matrix with some random cards from the baskets, so that the child had to tell the examiner that this was not the right answer.

While the first construction task (I) left freedom for the child to use any number and type of the 72 cards, the children were given a restricted set of cards for construction tasks II and III. The children were required to use all cards to make 'the puzzle'. The restricted set of cards guaranteed that in order to utilize all the given cards *and* construct a correct analogy, the task components number, colour, and size, (and animal for the 3rd task only) needed to be included. By their own volition, children could decide to make the constructed analogies even more complex by choosing to flip the cards and/or change the position.

For each of the tasks, the children were given as little instruction as possible in order to maximize spontaneous problem-solving behaviour. Some children, however, failed to start the task or forgot to leave one or more of the cells of the analogy open for the examiner to complete. In such

situations, the child was given up to a maximum of three small hints. Assistance was only given to help the child construct something that had the appearance of an analogy (with three filled cells and one empty cell) that the examiner could be asked to solve (see Appendix A for the procedure). After the child had finished creating the puzzle, the examiner placed a random animal card and asked a) if this was the correct answer, b) what was the child's correct answer, and c) and why that was the correct answer. All explanations given by the children about their analogies, including those that were provided before the examiner had asked for their explanations, were included in the scoring process.

Procedure

In the current study, the dynamic test design was changed into a microgenetic design by adding pre- and post-tests (pre-pre-training[yes/no]-post-post-transfer]. All children were seen individually, once a week, by the same examiner at each test session (seven, including the administration of Exclusion and the transfer session). At the start of each unguided practice 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 task-solving components used in the items. The examiner showed the animal cards and explained their features: colour, size, and the possibility to flip to the opposite direction. The examiner then turned to the first analogy item and stated that this was a 'kind of puzzle' with three boxes containing animals and a fourth empty box (C-term or D-term). The child was then asked 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´. The training and reversal (construction) procedures have been described in the paragraphs above.

Scoring and analyses

The outcome variables of the four unguided practice sessions were scored separately. Each child obtained, per session, a '*Complete analogies score'* (the sum of all analogies that were accurately solved, with a range from 0 to 20); a '*Task-solving components utilised score'* (the sum of accurately used task-solving components as evidenced by the child's behavioural solutions, with a range from 0 to 110); a *'Task-solving components explained score'* (the sum of all accurate verbally explained task-solving components following the question 'how did you solve the puzzle').

We aimed to assess the extent to which children's learning in relation to performance on a figural analogies task transferred to one that involved reversal. Several variables regarding children's performance on the figural analogies tasks of the four unguided practice sessions were used to predict children's performance on the transfer task. Progression scores of children's changing

performance on the three outcome variables of the unguided practice sessions were obtained by subtracting the percentages of session 4 (post-test) by the percentages of session 1 (pre-test) (see Appendix B). We discerned '*progression in accuracy*' (number of accurately solved analogies), '*progression in task-solving components utilised*' and '*progression in task-solving components explained*'.

We also included the quality of children's strategy-use at the four unguided practice sessions as a variable in some analyses. The following scoring system was used for each item: 1) *analogical strategy-use* (more than 50 percent of the task-solving components explained); 2) *partial analogical strategy-use* (less than 50 percent of the task-solving components explained, but at least one explained task-solving component), and 3) *inadequate, non-analogical strategy-use.* For each child we computed a sum score reflecting the percentage of the number of explanations that reflected partial analogical or full analogical strategy-use across all four unguided practice sessions. This variable will hereafter be called *overall strategy quality*. Progress in scores of strategy quality was calculated by subtracting the percentages of session 3 and 4 (post-tests) by the percentages of session 1 and 2 (pre-tests). This variable will hereafter be called *progression in strategy-use*.

Appendix B also provides the scoring system for the construction (reversal transfer) tasks. The first outcome variable of the construction tasks was the sum of accurately constructed analogies (range 0-3). This was an ordinal variable, thereby violating the assumptions of least-squares regression. We therefore performed ordinal logistic regression analysis (Agresti, 2007). The second outcome measure was the sum of the task-solving components that children used in their accurately constructed analogies. Outcome measure three was the sum of correct explanations for the task-solving components. These two variables were specified as counts (see Appendix B); we therefore used Poisson regression, and more specifically negative binomial regression analysis (Agresti, 2007). The following predictors were included in the regression analyses: condition, progress in analogical performance (consisting of progression in accuracy, task-solving components utilized and task-solving components explained), overall strategy quality, and strategy progression. Lastly, negative binomial regression analyses were conducted for each separate type of task-solving component. The regression analyses included the following variables: condition, overall strategy quality, and strategy progression.

Results

Before analysing our research questions, we conducted one-way ANOVAs to examine possible initial differences between the two treatment groups. The analysis did not reveal significant

difference between the children in the two groups regarding age (p= .72) or level of inductive reasoning (Exclusion) (p= .26). We also checked for possible differences in analogical reasoning performance at the first unguided practice session. Results showed no significant difference in number of complete, correct analogical solutions (p= .70), task-solving components (p= .45), and explained task-solving components (p= .44) between the two treatment groups.

Effects of unguided practice and training

We expected that performance on the figural analogies task as defined by children's complete analogies, use of task-solving components, and explanations of task-solving components would change as a consequence of practice alone but that trained children would outperform children in the unguided practice condition. Means and standard deviations for the three outcome variables are provided in Table 3. The changes over time were examined with repeated measures ANOVAs with Session (1 to 4) as within-subjects factor and Condition (unguided practice versus unguided practice plus training) as between-subjects factor. Results for the main effect of Session and the interaction effect of Session and Condition are shown in Table 4.

- Insert Table 3 -

Results for all three dependent variables (children's complete analogies; task solving components; explained task solving components) revealed a significant effect for Session (p< .001). Repeated contrast analyses for complete analogies showed a difference among the means for test sessions 1 and 2 (F(1, 102)= 40.58, p< .001, η_p^2= .29) and for test sessions 2 and 3 (F(1, 102)= 33.67, p< .001, η_p^2= .25). Apparently, children's accuracy changed as a consequence of unguided practice. As expected, the two conditions differed in degree of accuracy improvement, as revealed by a significant interaction effect of Session with Condition (p< .001). Subsequent within-subjects repeated-contrast analyses showed that the differentiation in performance accuracy between the two conditions took place between sessions 2 and 3 (F(1, 102)= 21.09, p< .001, η_p^2= .17). As shown in Table 3, both conditions improved their average number of complete analogies from session 2 to 3, but, as expected, the improvement was considerably larger for the children who were provided with training.

- Insert Table 4 -

Repeated contrast analyses for task components showed a difference among the means for sessions 2 and 3 ($F(1, 102)$= 32.00, p< .001, η_p^2= .24). Again, the two conditions differed in their improvement in their use of these across sessions, as shown by a significant interaction effect between Session and Condition (p< .001). Inspection of the results of within-subjects repeated-contrast analyses revealed differences between the conditions for sessions 2 and 3 ($F(1, 102)$= 16.40, p< .001, η_p^2= .14) and for sessions 3 and 4 ($F(1, 102)$= 5.11, p= .03, η_p^2= .05), with the trained children showing the most progression. Repeated contrast analyses for explained task-solving components revealed significant differences among the means for sessions 1 and 2 ($F(1, 102)$= 11.54, p= .001, η_p^2= .10), sessions 2 and 3 ($F(1, 102)$= 12.65, p= .001, η_p^2= .11), and sessions 3 and 4 ($F(1, 102)$= 6.34, p= .01, η_p^2= .06). These results indicate that, irrespective of condition, children improved their number of explained task-solving components as a consequence of repeated practice. As shown in Table 4, we again found a significant interaction effect between Session and Condition (p< .001). Within-subjects repeated-contrast analyses revealed differences for sessions 2 and 3 ($F(1, 102)$= 18.74, p< .001, η_p^2= .16) and 3 and 4 $F(1, 102)$= 4.26, p= .04, η_p^2= .04), indicating that the trained children showed significantly more improvements in the number of analogical explanations than children who were only provided with unguided practice.

We can conclude that the analogical reasoning performance of children changed as a consequence of their receiving opportunities for repeated unguided practice including or not including training. As expected, the training procedure embedded within dynamic testing, led to an improvement in analogical reasoning for all outcome measures greater than the effect of unguided practice alone. As expected, this was found for the number of analogies they accurately solved, their utilisation of task-solving components necessary to solve analogies, and their verbal explanations of the task-solving components they used to solve the analogy-problems.

Results from the role-reversal construction task

Next, we focused on children's performance regarding the construction tasks. We examined the number of accurately constructed analogies, the number of task components utilised, and their explanations of their actions. Firstly, we examined the children's overall ability to construct analogy-problems utilizing the role-reversal procedure. Fifty-six children were unable to construct any of the three analogies, with 27, 17, and 4 children constructing one, two, or three problems respectively. Although we thought that the first construction task, a fully open task with endless possibilities, would be the most difficult, it was found that the reduced option task was met with less success. Thus, 34 children were able to correctly construct an analogy-problem when options were fully open,

31 children succeeded in constructing the second task, and 8 children correctly produced a third task. Thus, at least one third of the children were more able to construct a correct analogy when they were given the freedom to use all of the 72 cards compared to construction tasks II and III where the children were given restricted sets of cards.

A logistic regression analysis was run to predict the number of accurately constructed analogies with condition (treatment group), progression in accuracy, overall strategy quality, and strategy progression as predictors. The outcomes (see Table 5) revealed that condition (p= .67), progression in accuracy *(p= .46)*, and progression in strategy-use (p= .27) did not significantly contribute to the prediction. Overall strategy quality, however, appeared to be a significant predictor (p= .01).

Negative binominal regression analyses (see Table 5) were run to predict both the number of task-solving components utilised by the child, and the number of task-solving components explained accurately after each construction task had been completed. On average the children included 3.37 (range 0-17) transformations (task-solving components) in their analogies. Again, children's overall strategy quality appeared to be a significant predictor (p<.01). No additional predictive value was found for condition (p= .87), progression in task-solving components utilised (p= .19), and progression in strategy-use (p= .25). Regarding the accurately explained task-solving components (transfer tasks); children mentioned, on average, 1.73 (range 0-10) components. The negative binomial regression analysis for this variable revealed no significant predictive contributions of condition and overall strategy quality (p= .57; p= .12). The progression in task-solving components explained (p= .04), and progression in strategy-use (p< .001), however, appeared to be significant predictors of the number of correctly explained analogies at transfer.

- Insert Table 5 –

In sum, contrary to our expectations, there was no effect of condition (unguided practice with or without training) for all outcome measures at transfer. There was some effect of overall strategy quality on children's reasoning accuracy and the number of task-solving components utilised in the constructed analogies. Finally, progression in children's explanations concerning task components and their progression in strategy-use was related to the number of correctly explained analogies at transfer.

Results per task component

As noted above, we investigated possible inter-individual differences in children's ability to correctly explain the role of task-solving components during the transfer task. We expected that the changes in colour, size, and quantity in the tasks children had constructed would be most frequently mentioned, with fewer references to the more challenging components, orientation and position. Table 6 provides an overview of the number of changes explained per task component type by both conditions. This shows that children from both conditions explained colour and size most frequently. Children in the unguided practice condition showed particular difficulty with position and orientation. Children in the training condition group explained more components overall and made more frequent references to orientation and position.

- Insert Table 6 -

We expected that trained children would mention all of the task-solving components more often than children in the unguided practice condition. Results using negative binomial regression analyses indicated that the differences between the two conditions for all task-solving components (animal, colour, size, number, position and orientation) were not significant ($p > .05$). Furthermore, the overall strategy quality of children during the unguided practice sessions was not related to the numbers of changes mentioned for the task components ($p > .05$). Progression in strategy-use was a significant predictor only for the task-solving component number ($\beta = -.08$, $p = .01$).

Discussion

A first objective of our study was to ascertain to what extent, and how, an elaborated form of dynamic testing based on graduated prompt techniques, could provide insight into children's potential for learning. Of particular interest were not only children's progression in solving figural open-ended analogy-problems but also the nature and variability of the changes in the strategies they employed, Our findings that the group that received training made greater progress than that which received practice opportunities alone corresponds with earlier studies, for example, by Carlson and Wiedl (1992) who reported that the most effective "dynamic" methods involved either elaborated feedback or overt, concurrent verbalisations.

One could argue the study lacked a control group of children who were provided solely with a pre-test and a post-test assessment without any feedback or explanation requested. Our design makes it difficult to answer the question whether children receiving unguided repeated practice (with verbal explanations) have shown greater progression in performance than would be achieved under strict control conditions. Other studies, however, that have used similar figural open-ended

analogies (Animalogica) with children of comparable age, have reported significant differences between these two groups of children (e.g., Stevenson et al., 2009, 2014, 2016).

As expected, children showed large variability in their ability to solve the analogy problems, the size of their gains, the number of prompts they needed, and their ability to explain why their answer was correct. These findings correspond with earlier literature reports (e.g., Resing & Elliott, 2011; Resing et al., 2015; Siegler & Svetina, 2002).

The problems children were asked to solve were tangible, rather difficult open-ended tasks. An important and intriguing advantage in utilising these kinds of materials and tasks lies in the fact that they offered us insight into underlying problem-solving processes (e.g., Harpaz-Itay et al., 2006; Resing & Elliott, 2011; Tzuriel & Galinka, 2000). The materials consisted of different task components (transformations), for instance, a large horse became a small horse, one big blue horse became two small yellow ones. As a result, the procedures employed were visible step-by-step. We knew, however, that such tasks would be particularly challenging and tailored assistance during the training phase would be necessary if at least some of the children were to be able to complete the transfer tasks (Stevenson et al., 2016). It therefore seemed justifiable to assume that the combination of dynamic testing and tasks requiring the construction of the answer would help to deepen children's understanding of the reasoning processes required for solving analogy problems.

A prominent second objective of our study was to see the extent to which the children would be able to transfer the newly learned knowledge and strategies to a very novel, but related, task. Why did we choose this transfer task? We assumed that our transfer task balanced between the two transfer mechanisms of "high road" and "low road" Salomon and Perkins (1989) discerned. On the one hand, children had to form an abstraction or scheme of the analogy-problems they were tackling during unguided practice and training sessions. On the other hand, the task format (matrix) and the very familiar materials (tangible cards) had a very close relationship with the tasks the children had practiced extensively at several occasions. The transfer task we used has to be considered as very new for the child, and required deep understanding of the process of solving analogies.

 One of the reasons for the unanticipated finding that performance on the transfer tasks was not superior for the trained group might be because these were too difficult to fully master in such a relatively short study period (see, for example, Bosma & Resing, 2008; Tzuriel & George, 2009), particularly for children of this age (Halford & McCredden, 1998). Another related reason might the rather short training that the children received. Many children managed to construct partially correct analogies. It was remarkable that the first open construction task, with "endless" possibilities, which we thought would prove to be the most difficult, was successfully completed by a third of the children. Their responses contained a large variety of transformations/task-solving components, and

17

up to ten analogical explanations of the problem-solving process. We noticed large individual differences in both treatment groups, and can conclude that at least some children of this young age have a sound understanding of how the process of solving figural analogies works.

As the dynamic-test-type training appeared to provide no additional improvement in transfer task performance over that of repeated unguided practice, it would appear that the training should be rendered more extensive, either by adding more items or increasing the number of training sessions (e.g. Tzuriel & George, 2009). Another suggestion for future research, supported by the observed differential patterns of progression in performance and strategy quality, is that children in this young age group might require a more tailored form of scaffolded instruction and feedback (Davidson & Sternberg, 1984; Siegler & Svetina, 2002), operating within their zone of proximal development (Alibali & Goldin-Meadow, 1993; Granott, 2002; Wood, Bruner, & Ross, 1976). The transfer tasks should have this open character with "endless" possibilities, giving children a chance to spontaneously create analogies at their level of performance/thinking, which could subsequently be evaluated on the basis of their complexity.

We also examined which variables predicted children's performance at the assessment phase of transfer. Children's overall strategy quality, reflecting the percentage of partial or full analogical strategy-use across all four testing sessions, appeared to be a significant predictor of both the number of the transfer tasks that the children were able to construct and the task components/ transformations their answers revealed. In addition, the quality of the children's explanations was predicted by the children's progress on the test items. It seems that both unguided practice and training influenced this explanation process: most probably, children had already learned to explain their answers during the test sessions, and were subsequently able to reverse roles. These findings correspond not only to Siegler's (2006) assumption that children will show transfer of knowledge to novel tasks once they have become good strategic solvers, but also to his, and Perry et al.'s (2010) finding that, before they are able to show transfer, children vary in their strategic behaviour. Other studies have also shown that high-level mastery in analogy performance is needed to detect transfer of learning at the deep transfer level we required from children (e.g., Day & Goldstone, 2012).

On the basis of our findings, we can conclude that the capacity to solve analogies is related to the capacity to construct them (see also Bosma & Resing, 2006; Harpaz-Itay et al., 2006). In accordance with our expectations, children were superior constructors of analogies when they executed their analogy-solving strategies (more) efficiently (Siegler, 2006). These outcomes offer a preliminary indication that those who consistently employed advanced solving strategies during the testing phase had acquired a deeper understanding of the underlying principles involved. After all, while constructing analogy tasks, children needed to extract the previously learned analogical

relationships from schemas in their memory, rather than working out existing relationships in the tasks presented to them (Harpaz-Itay et al., 2006; Martinez, 1999; Perkins, 1992).

At this point, one could wonder why our study did not include a baseline measure of each child's transfer abilities. We agree with other researchers (e.g., Barnett & Ceci, 2002; Perkins & Salomon, 1988; Salomon & Perkins, 1989) that content and context are crucial here. In addition, children have to be in a position to detect a rule, relationship or similarity between a base and a target problem. However, we do not consider it appropriate to have administered the transfer problems beforehand. In the training condition children explicitly learned how to find a rule to solve figural analogies. In this training condition and also the unguided practice condition, children made progress in this respect. The unique characteristic of our "reversal" transfer tasks lies in the fact that transfer was actively sought. Furthermore, the children were very young and we did not want them to become demotivated at an early stage of the study.

Although many children who did not fully completed their transfer tasks made reference to copying strategies, they also often included their own rules and, in contrast with their responses during the earlier assessment sessions, rarely reverted to idiosyncratic explanations such as "This little dog seems very lonely and therefore I gave him a companion". It is possible that multiple choice and even open-ended analogy task-formats encourage children to adopt strategies such as copying and storytelling (Martinez, 1999; Morrison, Doumas, & Richland, 2011; Stanger-Hall, 2012). In contrast, a more open transfer task, such as the one used in this study, may encourage the use of inductive reasoning and rule finding, a necessary component when solving analogies. Future studies should investigate whether rule-like solutions of this nature, derived from either dynamic testing or practice situations are able to provide additional data about the child's developing problem-solving capacities.

The current study has shown that an open construction task, serving as a measure of transfer, can provide additional information about young children's depth of learning and learning potential. Such information may prove to be of practical benefit to teachers, although more work is needed to justify such a claim. More specifically, this study suggests that knowledge of the types of strategies children utilise and verbalise can yield insights and understanding about children's cognitive potential. Such a conclusion has important implications for both individual and larger scale educational dynamic-test situations and particular curricular areas (e.g. Grigorenko, 2009; Haglund & Jeppsson, 2012), for example math and spelling. Whether analogy construction tasks provide more valuable information to educationalists when these are domain specific or domain general, such as the task reported in the present study, is a question that requires further investigation.

While dynamic testing is often presented as an interesting measure of one or more cognitive abilities, our findings show its potential for guiding differentiated forms of educational intervention. The procedure we utilised enables us to examine children's learning processes in depth, including progression in performance, types of prompts necessary for the particular child, progression in verbalisations, and changes in both strategy-use and strategy-quality from pre- to post-test. Certainly, for dynamic testing, related to the measurement of transfer, to inform educational practice adequately, it will be necessary to find ways to further integrate process-based assessment with classroom teaching. Approaches combining more extended, adaptive forms of dynamic testing, responsiveness to intervention, and curriculum-based intervention (e.g., Fuchs, Fuchs, Hollenbeck, 2007; Fuchs & Deshler, 2007; Grigorenko, 2009; Jeltova et al., 2011) seem to be the first necessary steps to further integrate assessment and intervention.

Although we must be careful about drawing conclusions from only one study, our evidence suggests that it is very difficult to achieve high-level transfer effects, based on the deep structure of a task and abstractions of both the problem-solving context and the types of tasks. Dynamic testing can result in significant gains but the effects appear to be restricted to the domain in which training took place and to tasks that are not very different from the original ones. Current procedures appear not to encourage children to spontaneously search for relational similarities between the new and previously learned tasks (e.g., Goswami, 1992; Opfer & Thompson, 2008). Future studies should stress this relationship more directly, for instance by using a second adaptive training procedure during the assessment of transfer.

In our opinion, process information of the kind provided in our study is both promising and challenging for educational and school psychologists. We also believe that our approach to studying transfer represents an intriguing new means to impact significantly upon both theory and practice (Elliott, 2003; Resing, 2013). Perhaps we should stress the importance of considering the variability within, and between, children's problem-solving processes. Too often it is assumed that "once learned is always learned" or "once accurate is always accurate" but as we have reported here, most children do not solve cognitive problems in this way; they experiment with different solving strategies (e.g., Siegler, 2007) and, in so doing, provide researchers with new complex problems to solve.

References

Agresti, A. (2007). *An introduction to categorical data analysis.* Hoboken, NJ: Wiley-Interscience.

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

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.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637.

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 & Child Psychology, 23*(3), 81–98.

Bosma, T., & Resing, W. C. M. (2008). Bridging the gap between diagnostic assessment and classroom practice. *Journal of Cognitive Education and Psychology, 7,* 174–196.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*, 61–100.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1, pp. 77–165). Hillsdale, NJ: Erlbaum.

Brown, A. L. (1982). Learning and development: The problems of compatibility, access and induction. *Human Development*, *25*(2), 89–115.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In P. Mussen (Ed.), *Handbook of Child Psychology* (Vol. 3, pp. 77–166). New York: Wiley.

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

Campione, J. C., Brown, A. L., Ferrara, R. A., Jones, R. S., & Steinberg, E. (1985). Breakdowns in flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence*, *9*(4), 297–315.

Carlson, J. S., & Wiedl, K. H. (1992). Principles of dynamic testing: The application of a specific model. *Learning and Individual Differences, 4*, 153–166.

Chen, Z., & Siegler, R. S. (2000). II. Overlapping Waves Theory. *Monographs of the Society for Research in Child Development*, *65*(2), 7–11.

Church, R. B. (1999). Using gesture and speech to capture transitions in learning. *Cognitive Development, 14,* 313–342.

De Corte, E. (2003). Transfer as the productive use of acquired knowledge, skills and motivations. *Current Directions in Psychological Science, 12*(4), 142–146.

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

Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology, 106*(2), 639–650.

Davidson, J. E., & Sternberg, R. J. (1984). The role of insight in intellectual giftedness. *Gifted Child Quarterly, 28*, 58–64.

Day, S. B., & Gentner, D. (2007). Nonintentional analogical inference in text comprehension. *Memory & Cognition*, *35*, 39–49.

Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist, 47*(3), 153–176.

Detterman, D. K. (1993). *Individual differences and cognition: Current topics in human intelligence.* Norwood, NJ: Ablex Publishing.

Elliott, J. G. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review, 55*(1), 15–32.

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

Engle, R. A. (2012). The resurgence of research into transfer: An introduction to the final articles of the transfer strand. *Journal of the Learning Sciences, 21*(3), 347–352.

Fabio, R. A. (2005). Dynamic assessment of intelligence is a better reply to adaptive behavior and cognitive plasticity. *The Journal of General Psychology*, *132*(1), 41–66.

Farrington-Flint, L., Coyne, E., Stiller, J., & Heath, E. (2008). Variability in children's early reading strategies. *Educational Psychology*, *28*(6), 643–661.

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

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*(2), 141–205.

Fuchs, D., & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice*, *22*(2), 129–136.

Fuchs, L. S., Fuchs, D., & Hollenbeck, K. N. (2007). Extending responsiveness to intervention to mathematics at first and third grades. *Learning Disabilities Research & Practice*, *22*(1), 13–24.

Goswami, U. (1992). *Analogical reasoning in children.* Hove, UK: Lauwrence Erlbaum Associates.

Granott, N. (2002). How microdevelopment creates macrodevelopment: Reiterated sequences, back-ward transitions, and the Zone of Current Development Zone. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 213–242). New York: 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.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124*(1), 75–111.

Haglund, J., & Jeppsson, F. (2012). Using self-generated analogies in teaching of thermodynamics. *Journal of Research in Science Teaching, 49*(7), 898–921.

Halford, G. S., & McCredden, J. E. (1998). Cognitive science questions for cognitive development: The concepts of learning, analogy, and capacity. *Learning and Instruction, 8*(4), 289–308.

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

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

Higgins, S. (2015). A recent history of teaching thinking. In R. Wegerif, l. Li, & J. C. Kaufman (Eds.), *The Routledge International Handbook of Research on Teaching Thinking* (pp. 19–28). New York, NY: Routledge.

Holyoak, K. J. (1984). Analogical thinking and human intelligence. *Advances in the Psychology of Human Intelligence, 2*, 199–230.

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.

Jacobs, P. I., & Vandeventer, M. (1971). The learning and transfer of double-classification skills: A replication and extension. *Journal of Experimental Child Psychology, 42*, 240–257.

Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko, E. L. (2011). Making instruction and assessment responsive to diverse students' progress: Group-administered

dynamic assessment in teaching mathematics. *Journal of Learning Disabilities, 44*(4), 381–395.

Kim, K., Bae, J., Nho, M. W., & Lee, C. H. (2011). How do experts and novices differ? Relation versus attribute and thinking versus feeling in language use. *Psychology of Aesthetics, Creativity, and the Arts, 5*(4), 379–388.

Klauer, K. J. (1992). Teaching inductive thinking to highly able children. *European Journal for High Ability*, *3*(2), 164–180.

Klauer, K. J. (1998). Inductive reasoning and fluid intelligence. A training approach. In W. Tomic & J. Kingma (Eds.), *Reflections on the concept of intelligence* (pp. 261–289). Greenwich,   CT: JAI Press.

Klauer, K. J. (2014). Training of inductive reasoning-An updated meta-analysis. *Zeitschrift Fur Padagogische Psychologie*, *28*(1-2), 5–19.

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

Kohnstamm, G. A. (1968). *Piaget's analysis of class inclusion: Right or wrong?* New York: Humanities Press.

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

Larsen-Freeman, D. (2013). Transfer of learning transformed. *Language Learning, 63*(1), 107–129.

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

Morrison, R. G., Doumas, L. A. A., & Richland, L. E. (2010). A computational account of children's analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science. 14*(3), 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.

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

Opfer, J. E., & Thompson, C. A. (2008). The trouble with transfer: Insights from microgenetic changes in the representation of numerical magnitude. *Child Development*, *79*, 790–806.

Pellegrino, J. W., & Glaser, R. (1982). Analyzing aptitudes for learning: Inductive reasoning. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 269–345). Hillsdale, NJ: Erlbaum.

Perkins, D. N. (1992). *Smart-schools: From training memories to educating minds.* New York, NY: Free Press.

Perkins, D. N., & Salomon, G. (1988). Teaching for transfer. *Educational leadership, 46*(1), 22–32.

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

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science, 21*(12), 1894–1902.

Pittman, K. M. (1999). Student-generated analogies: Another way of knowing? *Journal of Research in Science Teaching, 36*(1), 1–22.

Reeves, L. M., & Weisberg, R. W. (1993). On the concrete nature of human thinking: Content and context in analogical transfer. *Educational Psychology, 13*, 245–258.

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

Resing, W. C. M. (2000). Assessing the learning potential for inductive reasoning in young children. In C. S. Lidz & J. G. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications*. *Advances in cognition and educational practice (series editor: J.S. Carlson)* (Vol. 6, 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., Tunteler, E., & Elliott, J. G. (2015). The effect of dynamic testing with electronic prompts and scaffolds on children's inductive reasoning: A microgenetic study. *Journal of Cognitive Education and Psychology, 14*(2), 231–251.

Resing, W. C. M., Xenidou-Dervou, I., Steijn, W. M., & 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*(1), 144–150.

Salomon, G., & Perkins, D. N. (1998). Individual and social aspects of learning. *Review of Research in Education*, *23*, 1–24.

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., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General, 127*(4), 377–397.

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.

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

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. (2016). Dynamic resting of analogical reasoning in 5- to 6-Year-Olds: Multiple-choice versus constructed-response training Items. *Journal of Psychoeducational Assessment*. doi: 0.1177/0734282915622912

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

Stevenson, C. E., 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 & Child Psychology, 28*, 67–84.

Stanger-Hall, K. F. (2012). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE-Life Sciences Education*, *11*(3), 294–306.

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. (2002). Spontaneous analogical transfer in 4-year olds: A microgenetic study. *Journal of Experimental Child Psychology, 27*, 296–325*.*

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. (2013). Mediated learning experience strategies and cognitive modifiability. *Journal   of Cognitive Education and Psychology, 13*, 59–80.

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

Tzuriel, D., & George, T. (2009). Improvement of analogical reasoning and academic achievements by the Analogical Reasoning Program (ARP). *Educational and Child Psychology, 29*, 71–93.

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.

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.

Figure 1. Examples of figural analogies used during unguided practice and training sessions (adopted from Stevenson, Resing, & Froma, 2009).
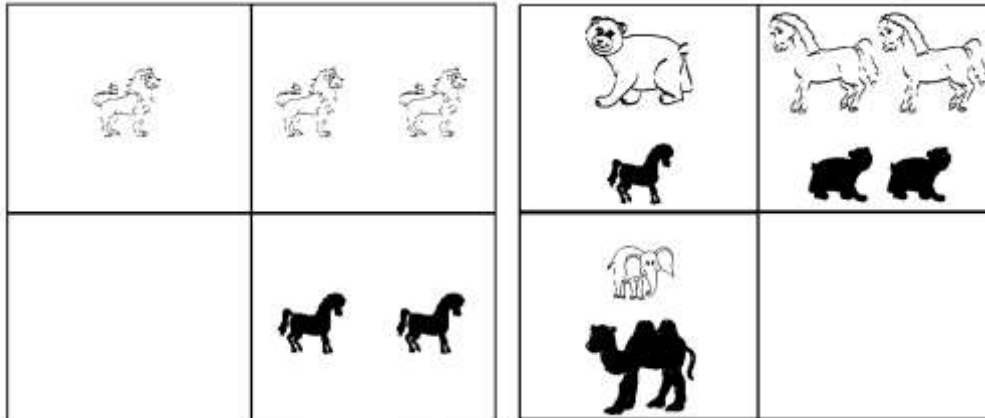
Note. Left figure: the lions are yellow; the horses are red. Right figure: the small horse, small bears and camel are blue; the large bear, large horses and elephant are yellow.

Table 1. Research design

Condition	Session						
	Exclusion	Pre-test 1	Pre-test 2	Training	Post-test 3	Post-test 4	Transfer
Practice	x	X	x	-	x	x	x
Training	x	X	x	x	x	x	x

Table 2. Schematic version of the graduated prompts procedure used during the training session

Type of prompt	Verbal instruction
Metacognitive	Experimenter asks the child to think about where to start and tells the child which boxes of the puzzle belong together.
Metacognitive	Experimenter explains that the boxes need to be compared and that the child needs to think about how the boxes belong together. The child can check if his or her answer is correct by comparing the boxes.
Task specific	Experimenter asks the child to explain changes from A to B and from A to C, and points out the similarity between [C:D] and [A:B], and between [B:D] and [A:C].
Task specific	Experimenter explains all task-solving components from A to B and from B to D
Task specific	Experimenter explains all task-solving components that occur in order to work towards the right solution

Table 3. Means and standard deviations of the analogical measurements for the practice (n = 52) and training (n = 52) conditions

Condition	Session 1	Session 2	Session 3	Session 4	Progress over time	Transfer session
	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
Complete analogies						
Practice	4.10(4.83)	5.58(5.54)	5.94(6.28)	6.31(6.10)	11.06(15.16)	0.67(0.88)
Training	4.44(4.33)	6.25(5.66)	9.38(5.35)	8.96(5.61)	22.60(20.73)	0.73(0.89)
Total	4.27(4.57)	5.91(5.58)	7.66(6.06)	7.63(5.98)	16.83(18.89)	0.70(0.88)
Task-solving components utilised						
Practice	51.44(31.00)	53.23(35.59)	56.31(37.22)	58.35(36.8)	6.28(11.88)	3.13(4.53)
Training	55.88(28.32)	58.52(33.02)	77.10(27.49)	76.08(26.73)	18.36(20.98)	3.60(4.54)
Total	53.66(29.63)	55.88(34.26)	66.70(34.19)	67.21(33.28)	12.32(18.02)	3.37(4.52)
Task-solving components explained						
Practice	20.29(21.11)	23.48(24.20)	22.54(24.24)	22.13(22.69)	0.81(10.55)	1.44(1.93)
Training	23.25(17.39)	26.98(21.38)	36.60(22.45)	32.52(23.51)	7.20(15.69)	2.02(2.42)
Total	21.77(19.31)	25.23(22.79)	29.57(24.24)	27.33(23.56)	4.00(13.68)	1.73(2.20)

Table 4. Results of the repeated measures ANOVAs for the number of complete analogies, task-solving components utilised, and explained task-solving components

	Wilks' λ	F	p	η_p^2
Complete analogies				
Session	.50	33.61	< .001	.50
Session x Condition	.81	7.85	< .001	.19
Task-solving components utilised				
Session	.68	15.42	< .001	.32
Session x Condition	.83	6.68	< .001	.17
Task-solving components explained				
Session	.77	9.93	< .001	.23
Session x Condition	.83	6.63	< .001	.17

Table 5. Results of the analyses for the number of correctly constructed analogies, task-solving components utilised, and task-solving components explained at transfer

	b(SE)	Exp(β)	Wald chi-square	p
Complete analogies[a]				
Condition	-.18(.43)	0.84	0.18	.67
Progression in accuracy	.01(.01)	1.01	0.56	.46
Overall strategy quality	.12(.05)	1.13	6.44	.01
Progression in strategy-use	-.02(.02)	0.98	1.22	.27
Task-solving components utilised[b]				
Condition	-.04(.26)	0.96	0.03	.87
Progression in task-solving components utilised	.01(.01)	1.01	1.72	.19
Overall strategy quality	.08(.03)	1.09	7.10	.01
Progression in strategy-use	-.02(.01)	0.99	1.32	.25
Task-solving components explained[b]				
Condition	.16(.27)	1.17	0.33	.57
Progression in task-solving components explained	.02(.01)	1.02	4.29	.04
Overall strategy quality	-.04(.03)	0.96	2.48	.12
Progression in strategy-use	-.05(.01)	0.95	13.36	< .001

Note. [a]Logistic regression; [b]Negative biniomial regression

Table 6. Number of explained changes for each task-solving component per condition at transfer

Condition	Animal	Color	Size	Number	Position	Orientation
Training	12	22	26	14	13	16
Practice	9	19	16	13	9	6
Total	21	41	42	27	22	22