The handle http://hdl.handle.net/1887/19813 holds various files of this Leiden University dissertation.

Author: Stevenson, Claire Elisabeth
Title: Puzzling with potential : dynamic testing of analogical reasoning in children
Issue Date: 2012-09-13
Prompting learning and transfer of analogical reasoning: Is working memory a piece of the puzzle?

This chapter is based on Stevenson, C. E., Heiser, W. J. & Resing, W. C. M. (under review). Prompting learning and transfer of analogical reasoning: Is working memory a piece of the puzzle?
4. **WORKING MEMORY AND TRANSFER OF ANALOGICAL REASONING**

**Abstract**

Dynamic testing is an assessment approach that aims to assess potential for learning by measuring performance improvement as a response to training while testing. In this study we use this approach in order to: (1) determine whether training children in analogical reasoning affects transfer of inductive reasoning skills to other tasks and (2) explore the relationship between working memory, training and transfer effects. This was investigated using a pretest-training-posttest control group design with 64 participants, aged 7-8 years (M = 7.6 years; SD = 4.7 months). All of the children were tested on four inductive reasoning tasks. Half of the children were trained in solving figural analogies according to the graduated prompts method, while the control group practiced with these items. Initial ability and performance change from pretest to posttest were estimated using Embretson’s (1991b) item response theory Multidimensional Rasch Model of Learning and Change. We found that the short training procedure improved figural analogical reasoning more than practice. Working memory was strongly related to initial performance on each of the inductive reasoning tasks. Yet, we found that performance change and knowledge transfer were only somewhat related to initial ability and unrelated to working memory. This indicates that performance change and ability to transfer trained skills to new tasks may be separate constructs and of possible importance in the assessment of learning and cognitive potential.

**Acknowledgments**

We would like to thank Hester van den Akker and Colette Kuipers for their assistance with data collection and coding and Hester for her additional contribution to data analysis.
4.1 Introduction

Dynamic testing can be defined as a diagnostic method that focuses on potential for learning and aims to provide insight into developing abilities (Elliott, 2003; Sternberg & Grigorenko, 2002). Dynamic assessment diverges from traditional assessment in that feedback is provided by the examiner during testing in order to facilitate learning and gain insight into learning efficiency and cognitive potential (Elliott et al., 2010). In dynamic testing, various indices are used to examine a child’s potential for learning, such as performance improvement following feedback interventions (e.g., Hessels, 2009; Tzuriel, 2001), the amount and type of instruction that best aids task solution (e.g., Bosma & Resing, 2012; Resing & Elliott, 2011), and ability to transfer these newly developed skills to other problems (Campione & Brown, 1987; Day et al., 1997; Lidz & Pena, 1996; Resing, 1997; Sternberg & Grigorenko, 2002). Previous research demonstrates that in dynamic testing designs using a pretest-training-posttest format the interventions generally lead to improve an examinee’s ability in the assessed skill (e.g., Day et al., 1997; Sternberg & Grigorenko, 2002). Furthermore, graduated prompting, a specific form of intervention, can provide insight into the examinee’s instructional needs (e.g., Bosma et al., submitted; Resing, Xenidou-Dervou, et al., 2012). In earlier dynamic testing research utilizing graduated prompting techniques the ability to transfer what was learned during the intervention was sometimes included in the assessment process (Brown & Kane, 1988; Campione, Brown, Ferrara, Jones, & Steinberg, 1985; Ferrara et al., 1986; Resing, 1993). However, transfer measures have received less attention in the more recent literature, perhaps due to the difficulty in eliciting transfer (e.g., Barnett & Ceci, 2002; Bransford & Schwartz, 1999; Detterman, 1993; Hager & Hasselhorn, 1998; Roth-Van Der Werf, Resing, & Slenders, 2002). Yet, transfer of skills to novel situations may provide insights into a child’s potential for learning (e.g., Bosma & Resing, 2006; Ferrara et al., 1986). In the present study we investigated the extent to
which reasoning skills learned during the dynamic testing of analogical reasoning were applied to similar untrained tasks. Furthermore, because inductive reasoning and working memory capacity (WMC) appear to be inter-related in children (e.g., Kail, 2007) we investigated the role of WMC on the near-transfer of inductive reasoning skills in a dynamic testing situation.

4.1.1 Dynamic testing of inductive reasoning

Dynamic tests often include inductive reasoning tasks (e.g., Ferrara et al., 1986; Resing & Elliott, 2011; Resing et al., 2009), which are considered central to intelligence (Carpenter et al., 1990; Klauer & Phye, 2008). Analogical reasoning, a form of inductive reasoning, is deemed essential to school learning and refers to the capacity to learn about a new situation by relating it to a structurally similar more familiar one (Goswami, 1992). Classical analogies (A:B::C:? and figural matrices (see Figure 4.1) are often included in measures of cognitive ability and considered strongly related to ‘g’ (Freund & Holling, 2011a; Primi, 2001). The ability to reason by analogy is assumed to develop with great variability throughout childhood (e.g., Leech et al., 2008; Siegler & Svetina, 2002; Tunteler & Resing, 2007a). Older children tend to perform better than younger children, which may be explained by improvements in efficiency of working memory capacity (Kail, 2007; Richland et al., 2006). Improvement in analogical reasoning can take place spontaneously with practice (e.g., Tunteler & Resing, 2002), with further learning effects provided by feedback (Cheshire et al., 2005), self-explanation (Siegler & Svetina, 2002; Stevenson et al., 2009) and other training formats (e.g., Alexander, Wilson, White, & Fuqua, 1987; Klauer & Phye, 2008; Tunteler et al., 2008). Training with graduated prompting techniques has been shown more effective than practice alone with regard to both learning and transfer (Bosma & Resing, 2006; Ferrara et al., 1986; Tunteler & Resing, 2010). Training type may also play a role in the learning and transfer of analogical
4.1. Introduction

reasoning (e.g., Harpaz-Itay et al., 2006; Stevenson, Heiser, & Resing, under review).

4.1.2 Transfer of inductive reasoning skills

The ability to spontaneously generalize a problem-solving approach taught in one context to a different but applicable situation is referred to as transfer. This is considered an important aim of formal schooling (e.g., De Corte, 2003). However, numerous studies show that transfer doesn’t occur easily as learning is context-bound and children rarely recognize that their acquired problem solving skills can be applied in novel situations (e.g., Barnett & Ceci, 2002; Bransford & Schwartz, 1999; Luo, Thompson, & Detterman, 2003; Siegler, 2006). Transfer can be assessed broadly such as from school learning to real-life situations or in a more narrow manner – from one cognitive task to a structurally similar one, referred to as near-transfer. Near-transfer also appears not to be common-place (see Jacobs and Vandeventer (1971) for this distinction). For example, Roth-Van der Werff et al. (2002) systematically assessed whether children trained in solving inductive reasoning tasks were able to generalize the learned problem solving skills to superficially similar and dissimilar problems that measured the same inductive reasoning skills. In their study, the children improved more on superficially similar tasks than those who only practiced with the same items. Yet, changes on superficially dissimilar tasks could be attributed to practice effects.

However, children may show greater transfer of knowledge when the targeted strategy has been mastered (Siegler, 2006). For example, Tunteler & Resing (2010) found that 8-year-olds who obtained high scores on a geometric analogy task improved more on a verbal analogies near-transfer task during the posttest. But as with Roth-Van der Werff et al. (2002) the improvement on the superficially dissimilar verbal analogy task in Tuntler & Resing’s study was independent of having received training – practice alone appeared to elicit transfer in high ability
4. **Working memory and transfer of analogical reasoning**

children. Aside from practice effects, instructional conditions also appear to play a role in near-transfer. For example, Harpaz-Itay, et al. (2006) found that 12-year-olds trained in verbal analogy solving also improved on geometric and numerical analogies, however the transfer effects were greater in children trained in analogy construction as opposed to multiple-choice solution.

In this study children were either trained to solve figural analogies in constructed-response format or practiced with these items (e.g., Stevenson et al., under review; Stevenson, Heiser, & Resing, submitted 2011a). We investigated whether training with graduated prompting or initial ability level played a role in the transfer of inductive solving skills to three related inductive reasoning tasks differing in content, format and/or measured construct. First, the geometric analogies used by Tunteler & Resing (2010), which differed only in content from the dynamically tested figural analogies. Second, an analogy construction task (e.g., Harpaz-Itay et al., 2006) in a form for younger children where roles of examiner and child are reversed (Bosma & Resing, 2006), which differed in format but not content or measured construct. Finally, a geometric and numerical seriation task (Durost, Gardner, & Madden, 1970), also included in Roth-Van der Werf et al.’s study (2002), was used that differed in content and construct (i.e. series completion rather than analogical reasoning).

4.1.3 **Working memory and inductive reasoning**

The influence of working memory capacity on the training and transfer of inductive reasoning in a dynamic testing context requires further research given that many researchers have found a strong relationship between working memory capacity (WMC) and inductive reasoning ability (e.g., Bacon, Handley, Dennis, & Newstead, 2008; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Krumm & Buehner, 2008; Kyllonen & Christal, 1990; Morrison, Holyoak, & Truong, 2001; Süb et al.,
Baddeley and Hitch (1974) proposed a model to describe the structure of working memory capacity (WMC) in which the central executive system is considered responsible for controlling attention and information processing, which regulates the operation of two domain-specific systems, the phonological loop and visuospatial sketchpad. The structure described by the Baddeley & Hitch model appears present and assessable in young children (Alloway, Gathercole, & Pickering, 2006; Gathercole, Pickering, & Wearing, 2004; Swanson, 2008) and related to young children’s analogical reasoning (e.g., Cho, Holyoak, & Cannon, 2007; Kail, 2007; Richland et al., 2006). For example, Krumm et al. (2008) found that working memory predicts a large amount of variance in reasoning ability. Furthermore, significant relations have been found between increases in efficiency of working memory capacity (WMC) and increases in reasoning and problem solving (Kail, 2007; Swanson, 2008). Tunteler & Resing (2010) found that memory of abstract figures was related to performance on the geometric analogies task, included as a transfer task in this study. Richland et al. (2006) found that children’s ability to solve scene analogies was related to their performance on a verbal WMC task. The separate contribution of the verbal and visuospatial components to figural analogy matrices utilized in the present study has not yet been investigated. We therefore extend the work of previous studies by including measures of both verbal and visuospatial WMC.

Working memory may become more efficient due to training and this automation of skills may result in greater transfer effects (Dahlin, Neely, Larsson, Bäckmann, & Nyberg, 2008; Jaeggi, Buschkuehl, & Perrig, 2008). It is plausible that training during dynamic assessment may lead to performance change and transfer effects through similar mechanisms. For example, we found that children’s WMC was related to improvement in analogy solving in untrained children but not in children who received training with graduated prompting techniques (Stevenson et al., submitted 2011a). Similarly, in a dynamic test utilizing the inductive reasoning task...
seriation, children with lower verbal WMC scores improved comparably to those with greater WMC scores but the gap was not closed (Resing, Xenidou-Dervou, et al., 2012). In Tunteler & Resing’s (Tunteler & Resing, 2010) microgenetic study including graduated prompting of geometric analogies the children with a less efficient WMC caught up with their peers with better WMC task performance after training. WMC appears related to training effects in dynamic tests. In the present study we broaden this investigation by examining whether the relationship of the dynamically assessed analogical reasoning skills and WMC extends to affect transfer to other inductive reasoning tasks.

4.1.4 Dynamic measurement of inductive reasoning

The dynamic test of figural analogies we administer contains a pretest, training and posttest. The outcomes of the dynamic test were pretest ability and performance change after training (posttest minus pretest) on the figural analogies task. In addition, we included the children’s performance change from pretest to posttest on geometric analogies and seriation transfer tasks, and their ability to solve an analogy construction transfer task administered only on the posttest. We look at change in performance over time, therefore it is important to pay attention to how we measure change because using classical test theory (CTT) scores, such as proportion correct, has received much criticism by psychometricians (e.g., Bereiter, 1963; Embretson, 1991b, 1991a; Lord, 1963; Prieler & Raven, 2002). The main problem with using CTT scores in a dynamic testing context is that when pretest and posttest scores are highly correlated, as is generally the case with repeated measures of the same construct, the change score is unreliable. This of course is unacceptable when one wants to reliably measure change. Furthermore, CTT scores are sensitive to bottom and ceiling effects and the meaning of change scores is dependent upon the examinee’s pretest performance. For example, an improvement of four correct solutions may
mean something different on a test of twenty items when the initial score was two or sixteen; a change in scores from two to six may represent greater improvement in understanding than sixteen to twenty. Item response theory (IRT), often referred to as modern test theory, offers solutions for the statistical pitfalls of measuring change with CTT e.g., Embretson & Reise, 2000. IRT scoring in its simplest form, the Rasch model, is based not only on the ability of the person taking the test, but also on the difficulty of the items included in the test. Embretson (1991b) proposed an IRT model, the Multidimensional Rasch Model for Learning and Change (MRMLC), that provides both reliable initial ability and change estimates that can be applied to dynamic testing (e.g., Dörfler, Golke, & Artelt, 2009; Embretson, 1987; Embretson & Prenovorst, 2000) and longitudinal research (e.g., Von Davier et al., 2010). We use this model for estimating the children’s pretest abilities and performance change from pretest to posttest.

4.1.5 Current study

In sum, this study investigated the effect of the graduated prompts training method on Rasch-scaled ability and performance change scores of figural analogies and inductive reasoning transfer tasks while examining the role of working memory capacity herein. In accordance with the literature described above we expected (hypothesis 1) the children’s performance change on the figural analogies task to be greater in children trained with graduated prompts than when only practicing with the items (see Stevenson et al., submitted 2011a). Transfer of reasoning skills was expected to coincide with initial ability (hypothesis 2a), where transfer effects would be greater in higher ability children (e.g., Tunteler & Resing, 2010). Trained children were expected to show greater transfer effects on the transfer tasks with differed only in content (geometric analogies) or format (analogy-construction) to the trained figural analogies task (hypothesis 2b: (e.g., Roth-Van Der Werf et al.,
4. WORKING MEMORY AND TRANSFER OF ANALOGICAL REASONING

Furthermore, we expected children with greater WMC to obtain to perform better on the figural analogies and transfer task pretests (hypothesis 3; e.g., Krumm & Buehner, 2008). The final aim was to explore the role of working memory in inductive reasoning transfer.

4.2 Method

4.2.1 Participants

Participants were 64 7-8 year olds (34 girls, 30 boys, M=7.6 years; SD=4.7 months). The children were recruited from three elementary schools located in two midsized towns in the Netherlands. The schools were selected based upon their willingness to participate. All children were native Dutch speakers. Written informed consent was obtained from the parents prior to participation.

4.2.2 Design & Procedure

A pretest-training-posttest control-group design with randomized blocking was employed. Children were blocked into a training or practice group for the ANIMALOGICA dynamic test based on scores on a visual exclusion test (Bleichrodt et al., 1987) and gender. Children were tested during six weekly sessions.

During the first session, the exclusion task and three working memory tasks were administered. In the next session all children were administered the figural analogies pretest. During the third session two transfer task pretests were administered: geometric analogies and seriation. The fourth session comprised of either training or practice in solving figural analogies. The fifth session consisted of the figural analogies posttest plus an analogy construction transfer task referred to as the reversal task. In the final session the geometric analogies and seriation transfer tasks were re-administered.
4.2. Method

**AnimaLogica**, the dynamic test of figural analogies used in this study, and working memory tasks were administered individually. Classroom-based administration was conducted for the exclusion task and the geometric analogies and seriation transfer tasks.

4.2.3 Instruments

**AnimaLogica: a dynamic test of figural analogical reasoning**

*Pretest and Posttest.* The figural analogies utilized colored (red, yellow or blue) animal figures, classically presented in 2x2 matrix format (e.g., Stevenson et al., 2009). Drawings of familiar animals occupied three squares and the lower left or right quadrant was empty. The transformations were made on the dimensions: (1) animal, (2) color, (3) size, (4) position, (5) orientation and (6) quantity. The child was asked to choose a picture from five options below to solve the puzzle (A:B::C:D). The five systematically constructed answer options included the correct answer, two partially correct answers (with one incorrect transformation) and two non-analogical answers (with 2 or more incorrect transformations). The test booklets each consisted of 30 items of increasing difficulty.

*Training.* The training items also consisted of figural analogy matrices. The objects and transformations were the same as the figural analogies task, but instead of multiple-choice items the training items were presented in constructed-response format (see Stevenson et al., under review). None of the 8 training items were identical to the test items. To solve the analogies, the children had to construct the solution from a number of animal cards representing the six transformations; each animal was available in three colors (red, yellow, blue), two possible sizes (large, small) and printed two-sided so by turning the card over the animal’s orientation could be changed (looking left or right). Quantity was specified by selecting one or
more animal cards and position was selected by placement in the empty box. An example item is shown in Figure 4.1.

Graduated prompting techniques (e.g., Campione & Brown, 1987; Resing, 1993; Resing & Elliott, 2011; Resing et al., 2009) were applied to aid the children in solving the training items. The stepwise instructions began with general, metacognitive prompts, such as focusing attention, followed by cognitive hints, emphasizing the transformations and solution procedure, and ended with step-by-step scaffolds to solve the problem (see Stevenson et al., submitted 2011a). A total of five prompts were administered. During the first prompt the child was asked how such an item was solved previously and was provided with more detailed instruction, thereby aiding problem recognition and redefinition. During the second prompt a card was presented which included the general steps to solve the analogies, analogous to Sternberg’s (1977) componential analysis: (1) look closely (encoding component), (2)
think about how the animals change (inference component), (3) apply this to solve for the empty box (mapping component) and (4) check your work (justification component). In the third prompt, these components were further emphasized while the examiner worked through the steps on the aid card with the child, explaining with both words and gestures. For example, “What changes from here to here (A:B)?”. In the fourth prompt the horizontal and vertical transformations were summarized, emphasizing inference and encouraging mapping. In the final prompt the examiner used scaffolds to help the child systematically solve the problem per transformation, such as “Which animals belong in the empty box?”, “Which direction should the dog face?”. After each question direct feedback was given, guiding the child step-by-step to the correct solution. Once the child answered an item correctly the child was asked to explain his/her answer; no further prompts were provided and the examiner proceeded with the next item.

Transfer tasks

The three transfer tasks were selected because each has been used in previous studies assessing inductive reasoning transfer in children. Each task differed from the main task, figural analogies, with regard to content, format and/or measured construct.

*Geometric analogies.* The geometric analogy task (Hosenfeld & Resing, 1997) consisted of 20 multiple-choice items (see Figure 4.2). The child had to choose the correct answer from five options. The content differs from the figural analogies in that geometric objects instead of animal figures are used. Otherwise the tasks are superficially similar and both require analogical reasoning skills and are presented in multiple-choice format.

*Seriation.* The seriation task (Durost et al., 1970) consisted of 20 numerical and 14
4. Working memory and transfer of analogical reasoning

Figure 4.2 An example item from the geometric analogies transfer task (Hosenfeld et al., 1997).

geometrical seriation items respectively (see Figure 4.3). The answer to complete the series is selected from four or five options on the right-hand side of the row. The content, geometric objects, is similar to that of the geometric analogies, but different from the animal figures in the figural analogies. This task is also presented in multiple-choice format but requires a different form of inductive reasoning than the figural analogies task, namely series completion, and therefore differs in the measured construct.

Reversal Task. The reversal task is an analogy construction task in which the child is asked to take on the role of teacher (Bosma & Resing, 2006) and construct a matrix analogy for the examiner. The content of this task is the same as the figural analogies task as the same animal figures were used as in the ANIMALOGICA task, but here the matrix was empty (see Figure 4.4. The format was different because
4.2. Method

Figure 4.3 Three example items from the seriation transfer task (Durost, et al., 1970).

The child was asked to construct an analogy and instruct the experimenter on how to solve it. This task requires understanding of analogical reasoning to be able to construct a correct analogy (e.g., Harpaz-Itay et al., 2006).

Working memory

Backward Digit Span. The WISC IV Digit Span backwards (Wechsler, 2003) is considered a measure of verbal working memory capacity (e.g., Süb et al., 2002). The child is asked to repeat a sequence of digits in reverse order.

Listening Recall. The Automated Working Memory Assessment (AWMA, Alloway, 2007) listening recall consists of spoken sentences, of which the child is asked to repeat the first word and say whether the sentence is true or false (e.g., bicycles can walk). This task measures verbal working memory.

Spatial Span. In the AWMA (Alloway, 2007) spatial span subtest, a sequence of two figures are presented and the child is asked to say whether these are the same or different. In some cases one of the figures is rotated (i.e. same) and others mirrored.
and rotated (i.e. different). The child must also recall in sequence whether the red dots were located above, left or right of the figure on the right. This task measures visuospatial working memory.

**Visual exclusion**

Visual exclusion is a subtest of the Revised Amsterdam Children’s Intelligence Test (RAKIT, Bleichrodt et al., 1987) used to measure visual-spatial inductive reasoning ability. The child is shown four abstract geometric figures and asked to choose which one doesn’t belong to the other three.
4.3 Results

4.2.4 Scoring

The children’s answers to the figural analogies, geometric analogies and seriation items were based on the selected or constructed answer and scored as correct/incorrect (skipped items were scored as incorrect). Rasch estimates from item response theory (IRT, e.g., Embretson & Reise, 2000) were obtained for the initial ability (pretest performance) and performance change (gain from pretest to posttest) using Embretson’s Multidimensional Rasch Model for Learning and Change (MRMLC, Embretson, 1991b, 1991a). Initial analyses were conducted with the ltm package for R (Rizopoulos, 2006) and the MRMLC estimates were computed using the lmer4 package (Bates & Maechler, 2010).

The measure of children’s performance on the reversal items was based on a combination of whether they could correctly construct an analogy and the complexity of the analogy, represented by the number of transformations present (e.g., Hosenfeld & Resing, 1997; Mulholland et al., 1980; Stevenson et al., 2009, 2011). The resulting score on this analogy construction task was correctness (1/0) x number of represented transformations (1-6).

4.3 Results

Before conducting analyses to answer the research questions we first describe the psychometric properties of the Rasch-scaled tests and items. Furthermore, we check whether the children in the two conditions differed in cognitive functioning or age prior to testing.

4.3.1 Psychometric Properties

Pretests and posttests were administered for the figural analogies (FA), geometric analogies (GA) and seriation (SR) tasks. Cronbach’s measure of internal consistency
4. WORKING MEMORY AND TRANSFER OF ANALOGICAL REASONING

on the pretests was $\alpha = .81$, $\alpha = .92$ and $\alpha = .91$ for FA, GA and SR tasks respectively. For the posttests this was $\alpha = .83$, $\alpha = .92$ and $\alpha = .91$ for the FA, GA and SR tasks.

Before applying the MRMLC model, first the independent Rasch model parameters were estimated for the pretests and posttests using marginal maximum likelihood (MML) estimation. The parametric Bootstrap goodness-of-fit test in the ltm package was used to investigate model fit. The Pearson’s $\chi^2$ statistic (based on a comparison with 50 generated datasets) indicated that the Rasch model fit of the pretests and posttests were acceptable ($p > .05$) with the exception of the seriation pretest ($p = .02$). The correlations between the item parameters of the pretests and posttests were very strong for each of the tasks, $r_{FA} = .76$, $r_{GA} = .87$ and $r_{SR} = .91$, therefore we considered the application of Embretson’s MRMLC model appropriate. The range of the MRMLC item difficulty parameters was $-2.60$ to $3.40$ ($M = .83, SD = 1.60$) for the FA task, $-3.03$ to $1.33$ ($M = -.20, SD = 1.02$) for GA task and $-4.42$ to $1.69$ ($M = -.75, SD = 1.18$) for the SR task.

4.3.2 Initial Group Comparisons

The children’s average age ($F(1, 62) = 2.13, p = .15$), initial level of inductive reasoning (visual exclusion: $F(1, 62) = .27, p = .61$), working memory capacity (backward digit span (BDS): $F(1, 62) = .23, p = .64$; listening recall (LR): $F(1, 62) = .02, p = .88$; spatial span (SS): $F(1, 62) = .48, p = .49$) and pretests (figural analogies: $F(1, 62) = .36, p = .55$; seriation: $F(1, 62) = .00, p = .98$) did not differ between conditions (see Table 4.1 for basic statistics). Initial performance on geometric analogies pretest differed significantly between conditions: $F(1, 62) = 5.45, p = .02$.

4.3.3 Effect of graduated prompting on figural analogy solving

Our first research question concerned the effect of the graduated prompts training in improving the children’s performance on the figural analogies task. We expected
4.3. Results

Table 4.1 Basic statistics of age, exclusion, working memory and pretest scores (MRMLC ability estimates) of figural analogies, geometric analogies and seriation per condition.

<table>
<thead>
<tr>
<th></th>
<th>Training (N=32)</th>
<th>Practice (N=32)</th>
<th>Total (N=64)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Age</td>
<td>92.38</td>
<td>5.12</td>
<td>90.66</td>
</tr>
<tr>
<td>Visual exclusion</td>
<td>16.06</td>
<td>6.92</td>
<td>16.97</td>
</tr>
<tr>
<td>Working memory:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit span backwards</td>
<td>6.71</td>
<td>1.90</td>
<td>6.94</td>
</tr>
<tr>
<td>Listening recall</td>
<td>12.58</td>
<td>3.75</td>
<td>12.71</td>
</tr>
<tr>
<td>Spatial span</td>
<td>16.48</td>
<td>6.41</td>
<td>17.55</td>
</tr>
<tr>
<td>Pretest score:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figural analogies</td>
<td>-.0772</td>
<td>.7717</td>
<td>-.0591</td>
</tr>
<tr>
<td>Geometric analogies</td>
<td>-.4604</td>
<td>1.4767</td>
<td>.4415</td>
</tr>
<tr>
<td>Seriation</td>
<td>-.0203</td>
<td>1.2422</td>
<td>-.0118</td>
</tr>
</tbody>
</table>

(1) that graduated prompts techniques would lead to greater improvement in analogical reasoning scores than practice alone. This was investigated using an analysis of variance (ANOVA) with figural analogy performance change estimates as the dependent variable and pretest score as the covariate. There was a between-subjects effect for condition ($F(1, 61) = 3.99, p = .05, \eta^2_p = .06$) indicating that the conditions differed in their degree of improvement. The covariate, pretest score, did not affect the change score: $F(1, 61) = 1.14, p = .29, \eta^2_p = .02$. Inspection of the means and standard deviations (see Table 4.3.3) shows that the children in the training condition obtained significantly higher performance change scores than those in the practice condition, confirming hypothesis 1.
4. Working memory and transfer of analogical reasoning

Table 4.2 Basic statistics of performance change from pretest to posttest (MRMLC) and reversal task performance.

<table>
<thead>
<tr>
<th></th>
<th>Training (N=32)</th>
<th>Practice (N=32)</th>
<th>Total (N=64)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Change scores:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figural analogies</td>
<td>.0673</td>
<td>.2683</td>
<td>-.0808</td>
</tr>
<tr>
<td>Geometric analogies</td>
<td>-.0199</td>
<td>.4995</td>
<td>-.0192</td>
</tr>
<tr>
<td>Seriation</td>
<td>-.0081</td>
<td>.8012</td>
<td>.0007</td>
</tr>
<tr>
<td>Reversal task score:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analogy Construction</td>
<td>2.47</td>
<td>2.19</td>
<td>1.69</td>
</tr>
</tbody>
</table>

4.3.4 Effect of graduated prompting on transfer

The second research question related to the children’s ability to transfer learned figural analogical reasoning skills to geometric analogies (GA), seriation (SR) and an analogy construction (AC) task. We expected (2a) transfer effects to be related to the children’s pretest scores and expected (2b) to find a training effect on the transfer tasks of similar content or format (GA and AC). For the GA and SR tasks transfer was ascertained using the change score – i.e. degree of improvement from pretest to posttest. For the AC task, which was not pretested, the reversal task score (see 4.2.4 Scoring) was used as a transfer measure.

Before testing our hypotheses we computed the correlations between the FA pretest scores and performance on the three transfer tasks. The FA pretest score correlated strongly with the pretest scores on the GA ($r = .57, p < .001$) and SR tasks ($r = .63, p < .001$). The AC score was moderately correlated with the FA pretest score ($r = .37, p = .003$).

To investigate the relationship of transfer with condition and FA pretest performance a MANCOVA (3 transfer tasks x 2 conditions) with GA change...
estimates, SR change estimates and AC scores as dependent variables with figural analogy pretest score as covariate was conducted. An effect was found for FA pretest score on AC performance ($F(1, 61) = 9.33, p = .003, \eta^2_p = .13$), but not on the change scores of GA ($F(1, 61) = .11, p = .74, \eta^2_p = .00$) or SR ($F(1, 61) = 1.71, p = .20, \eta^2_p = .03$). Transfer effects appear only partially related to FA pretest scores (Wilks’ $\lambda = .85, F(3, 59) = 3.44, p = .02, \eta^2_p = .15$); hypothesis 2a is only partly accepted. Results show that condition does not lead to a differential effect on transfer (Wilks’ $\lambda = .96, F(3, 59) = .79, p = .51, \eta^2_p = .04$), hypothesis 2b is rejected.

4.3.5 Role of working memory in analogical reasoning ability and transfer

Our third research question pertains to the role of working memory in analogical reasoning ability and transfer. We expected (3a) working memory capacity to be related to the children’s performance on all tasks. We also explored whether (3b) WMC was positively related to transfer effects.

First correlations were used to examine the relation between WMC and children’s performance and change on all tasks. Backward digit span (BDS) showed a moderate correlation with FA, GA and SR (see Table 4.3.5), confirming hypothesis 3a. The correlations of the change estimates and AC scores with WMC were not significant; therefore hypothesis 3b could be rejected.

The three working memory tasks were not strongly correlated (BDS, LR $r = .10, p = .46$; BDS, SS $r = .34, p = .01$ and LR, SS $r = .34, p = .07$). In order to gain greater insight into the WM components involved in each of our experimental tasks, we further investigated whether combinations of the working memory measures explained significantly greater variance in pretest scores than just BDS. Hierarchical regression analyses were conducted with BDS entered as the first predictor and LR or SS as the second variable. In the case of figural analogies the best fitting model included listening recall ($\Delta R^2 = .061$) in addition to BDS explaining 16.6% of the
4. Working memory and transfer of analogical reasoning

Table 4.3 Correlations of working memory measures and pretest and change scores of figural analogies, geometric analogies, seriation and reversal analogy construction score.

<table>
<thead>
<tr>
<th></th>
<th>Backward digit span</th>
<th>Listening Recall</th>
<th>Spatial Span</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretest score:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figural analogies</td>
<td>.323*</td>
<td>.277*</td>
<td>.268*</td>
</tr>
<tr>
<td>Geometric analogies</td>
<td>.347*</td>
<td>.101</td>
<td>.303*</td>
</tr>
<tr>
<td>Seriation</td>
<td>.385*</td>
<td>.206</td>
<td>.345*</td>
</tr>
<tr>
<td><strong>Change score:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figural analogies</td>
<td>-.003</td>
<td>-.081</td>
<td>.145</td>
</tr>
<tr>
<td>Geometric analogies</td>
<td>.063</td>
<td>-.168</td>
<td>.044</td>
</tr>
<tr>
<td>Seriation</td>
<td>.113</td>
<td>.012</td>
<td>.069</td>
</tr>
<tr>
<td><strong>Reversal task score:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analogy Construction</td>
<td>.104</td>
<td>.215</td>
<td>-.041</td>
</tr>
</tbody>
</table>

*p < .05

variance (see Table 4.3.5). For GA and SR, neither LR or SS explained significant additional variance, although in both cases when BDS was excluded from the analyses SS was the best predictor (see Table 4.3.5).

4.4 Discussion

The main aim of this study was to investigate the learning and transfer of analogical reasoning skills in a dynamic testing context and explore the role of working memory capacity herein. We compared the learning and transfer of inductive reasoning skills of children who were trained during dynamic testing with graduated prompts or practiced without feedback on a figural analogies task. As with previous studies (e.g., Siegler & Svetina, 2002; Tunteler & Resing, 2010), we found that trained children showed greater progression in analogy solving than children in the practice condition. Furthermore, performance on the figural analogy matrices
4.4. Discussion

Table 4.4 Results of hierarchal linear regression analyses predicting pretest scores from working memory measures.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Predictor</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Figural analogies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td>Backward digit span</td>
<td>.148</td>
<td>.059</td>
<td>.300*</td>
</tr>
<tr>
<td>Step 2</td>
<td>Listening recall</td>
<td>.070</td>
<td>.033</td>
<td>.249*</td>
</tr>
<tr>
<td><strong>Geometric analogies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>Backward digit span</td>
<td>.299</td>
<td>.104</td>
<td>.347**</td>
</tr>
<tr>
<td>Model 2</td>
<td>Spatial span</td>
<td>.071</td>
<td>.025</td>
<td>.345**</td>
</tr>
<tr>
<td><strong>Seriation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>Backward digit span</td>
<td>.259</td>
<td>.080</td>
<td>.385**</td>
</tr>
<tr>
<td>Model 2</td>
<td>Spatial span</td>
<td>.080</td>
<td>.032</td>
<td>.303*</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

Pretest was strongly related to performance on each of the transfer tasks: geometric analogies, seriation and analogy construction. This coincides with previous research as the relationship between these tasks has been emphasized in numerous studies (e.g., Carpenter et al., 1990; Roth-Van Der Werf et al., 2002; Sternberg & Gardner, 1983). Transfer of analogical reasoning skills to the reversal situation in which the child constructed an analogy for the examiner was related to initial ability on the figural analogies tasks, where more complex analogies were constructed by the children with higher pretest scores. The findings on the reversal task were in line with Siegler’s theory (Siegler, 2006) that greater mastery of task strategies increases the chances of knowledge transfer to a novel situation in children. Yet as with previous research on the effect of the short graduated prompts training on transfer of inductive reasoning skills (e.g., Tunteler & Resing, 2010; Roth-Van Der Werf et al., 2002), we found that children in the training condition showed a similar degree of improvement from pretest to posttest on transfer tasks with dissimilar content.
4. **Working memory and transfer of analogical reasoning**

as the children who only practiced with the items. A possible explanation for our results and those of previous studies where training on a different task does not affect transfer of knowledge to similar tasks stems from Opfer & Thompson’s (2008) practice interference hypothesis. Their theory suggests that practice using incorrect solution strategies, which often occurs during pretesting, which was included in the present and previous studies, impedes transfer. This could explain why transfer effects were only found on the reversal task which was not pretested, but not on the other two transfer tasks in this study. In the assessment of transfer within dynamic tests, which often comprise a pretest-training-posttest format, it is perhaps advisable not to pretest the transfer tasks. Instead a selection of transfer tasks that measure similar skills to the tested task may provide more reliable measures. The effect of initial ability could be accounted for using the pretest scores of the trained task, which indeed correlated with performance on the analogy construction (reversal) task in the present study. However, greater transfer of knowledge has been demonstrated in other research in which pretests were included but more extensive training was provided (e.g., Harpaz-Itay et al., 2006; Klauer & Phye, 2008; Rittle-Johnson, 2006; Siegler & Svetina, 2002). We therefore advise including more training sessions to investigate whether trained children would show greater transfer on a group level than practice or control groups to further verify the effects of the graduated prompting procedure.

The goal of dynamic testing, however, is to ascertain the amount of learning and transfer an individual can achieve after a short training procedure in order to gain insight into learning efficiency. In order to assess this we used item response theory (IRT) Rasch estimates of the degree of performance change the children showed from pretest to posttest. These estimates provide a more accurate picture of proficiency change by avoiding statistical pitfalls of traditional scores, such as percentage correct, where change scores are unreliable and bottom or ceiling effects...
4.4. Discussion

could warp the degree of performance change (e.g., Bereiter, 1963; Prieler & Raven, 2002). We used Embretson’s (1991b, 1991a) MRMLC model which provides reliable change estimates to measure training and transfer effects on the pretested tasks.

Our results show great variability in initial ability and performance change on of each of the inductive reasoning tasks and we therefore investigated whether working memory capacity could be a source of individual differences.

A great deal of research with adults has demonstrated that working memory capacity is strongly related to fluid intelligence and inductive reasoning (e.g., Ackerman, Beier, & Boyle, 2005). It is also postulated to be a bottleneck in children’s analogical reasoning (Richland et al., 2006; Thibaut, French, & Vezneva, 2010). Our results coincide with this as we found moderate correlations between working memory measures and the children’s initial ability levels on all three inductive reasoning tasks. This relationship \( r \approx 0.35 \) was not as strong as in adult populations, but similar to that found in other research with children (Alloway, Gathercole, Willis, & Adams, 2004; Hornung, Brunner, Rueter, & Martin, 2011; Tillman, Nyberg, & Bohlin, 2008). Verbal WM played a stronger role in the solution of figural analogies and visuo-spatial WM contributed more to performance on the geometric analogies and the geometric and numerical seriation task. These findings are in line with Hornung et al. (2011) where substantial relationships were found between the verbal and visuospatial WM factors with young children’s performance on Raven’s colored matrices – a task which among other traits also requires inductive reasoning to solve. However, given their conclusion that short-term memory best explains the relationship between working memory and Raven performance, it is advisable to include short-term memory in future investigations of the role of memory in children’s performance on inductive reasoning tasks.

From the literature and our results we can conclude that WMC is related - to a certain degree - to inductive reasoning ability in children (Engel de Abreu, Conway,
4. Working memory and transfer of analogical reasoning

& Gathercole, 2010; Hornung et al., 2011; Richland et al., 2006; Tillman et al.,
2008; Tunteler et al., 2008; Tunteler & Resing, 2010). Given the importance placed
upon WMC in cognitive and psychoeducational assessment (Hatcher, Snowling, &
Griffiths, 2002; Martinussen, Hayden, Hogg-Johnson, & Tannock, 2005; Pickering
& Gathercole, 2004) the question arises whether WMC can explain individual
differences in the amount of learning and transfer a child demonstrates in a
dynamic assessment procedure. In this study, we found WMC was unrelated to
the children’s improvement on the trained task or degree of transfer to related
tasks after training. It appears that WMC does not sufficiently explain individual
differences in learning or transfer in a dynamic testing context. Our analysis of the
role of WMC was exploratory and the study comprised of a small sample, therefore
more extensive research is needed to substantiate our findings.

Inductive reasoning ability and WMC are well-established constructs in cognitive
ability tests and known to be related. Performance change and ability to transfer
knowledge to novel situations, such as in the reversal task, are less often included
in the assessment of intellectual abilities (Bosma & Resing, 2006; Elliott et al.,
2010). Our finding that change scores and knowledge transfer are only somewhat
related to initial ability and unrelated to WMC indicates that these may be separate
constructs and important in the assessment of learning and cognitive potential.
Further research should focus on the relevance of change scores and performance on
transfer tasks in psychoeducational assessment – whether these constructs provide
a better picture of a child’s capabilities and potential.