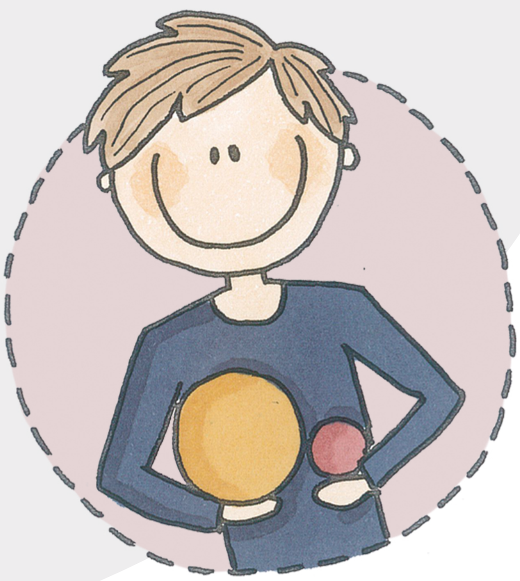
Computerised dynamic testing: An assessment approach that tailors to children's instructional needs

Touw, K.W.J.

Author: Touw, K.W.J.
Title: Computerised Dynamic Testing: An assessment approach that tailors to children's instructional needs
Issue Date: 2020-09-17

Chapter 4

Dynamic testing of children with a developmental language disorder: Progression in series completion performance and need for instruction



Kirsten W. J. Touw
Bart Vogelaar
Floor Thissen
Wilma C. M. Resing

Touw, K. W. J., Vogelaar, B., Thissen, F., & Resing, W. C. M. (2019). Dynamic testing of children with a developmental language disorder: Progression in series completion performance and need for instruction *Educational and Child Psychology, 36*(1), 27-39

Abstract

Background: Outcomes of static tests provide an indication of what children have learned in the past, up to the moment of testing, and can therefore underestimate the cognitive abilities of atypically developing children, such as children with language difficulties. In contrast, dynamic tests aim to examine children's potential for learning. The information obtained during dynamic testing has the potential to be helpful in education. However, the administration of dynamic tests is often time consuming.

Aim: The current study aimed to investigate the value of a computerised dynamic test utilising graduated prompts, as a means to provide insight into the extent to which children with language difficulties show progression in accuracy of their series completion skills (a subform of inductive reasoning). We sought to investigate whether dynamically versus statically tested children diagnosed with a Developmental Language Disorder (DLD) attending special education showed differential patterns in their serial reasoning performance and their need for instruction.

Rationale/Approach: Forty seven children with a mean age of 8.06 years participated in the study. They were tested dynamically by means of a pre-test, two training sessions, and a post-test. We examined in detail children's learning processes when solving visual-spatial series completion tasks.

Findings/Conclusion: The graduated prompts training resulted in larger changes in accurately applied transformations than repeated practice only. Furthermore, children with weaker initial scores seemed to show larger progression paths after training than those with stronger initial scores, although no significant differences were found. The current study also reveals that, in general, children needed fewer prompts in the second training session than in the first, although large individual differences in the amount and type of help needed were observed. Overall, these findings provide a first indication of the value of a computerised dynamic test in unveiling the potential for learning and individuals' needs for instruction in the population of children with DLD.

4.1 Introduction

Dynamic testing can be seen as a process-oriented assessment technique, in which instruction and feedback are interwoven in the testing process (Sternberg & Grigorenko, 2002). The feedback given during the testing process has been said to provide insight into the manner in which a child learns (Bosma & Resing, 2012; Resing & Elliott, 2011). In contrast, static tests are characterised by single moment test administration, reflecting only what a child has learned up to that time, and do not provide information regarding the ability to learn (Elliott et al., 2010). Static test scores may, therefore, lead to an underestimation of children's cognitive abilities, for example, for children that have language disorders (e.g., Camilleri & Botting, 2013; Elliott et al., 2010; Hasson & Botting, 2010; Poehner & Lantolf, 2005, 2013). This leaves educational professionals working with children with language disorders with additional questions, such as 'is this the child's actual level, or is the child capable of more than he or she has shown so far?' As dynamic testing is assumed to reveal more information regarding children's learning processes and cognitive potential, the current study aimed to investigate the value of a computerised dynamic test as a means to provide insight into the cognitive abilities of children with a developmental language disorder (DLD).

Developmental language disorder (DLD)

Over the past decades, numerous definitions of language disorders have been presented. These definitions often contain general descriptions like 'language skills that are different from the average'. In the Netherlands, in both healthcare and education, consensus has been reached on using the term 'developmental language disorder' (*Taalontwikkelingsstoornis,* in Dutch) when referring to children with said language difficulties (Gerrits et al., 2017). Recently, Bishop et al. (2017) published a study on the consensus on the English term for a language disorder. She also opted for 'developmental language disorder' (DLD) instead of 'specific language impairment' (SLI). Children with DLD constitute a heterogeneous group, and several subgroups can be distinguished, such as children with expressive language disorder, mixed receptive-expressive language disorder, and/or phonological disorder. In the literature, the reported prevalence of children with DLD ranges from five percent to 12 percent in children aged 0-7 years (Law et al., 2000; Shriberg et al., 1999). A frequently cited prevalence rate is the seven percent from the study by Tomblin et al. (1997). Developmental language disorders are more commonly found in boys than in girls; in clinical practice, the ratio is 3:1. Several studies focusing on young children's language and cognitive development have shown an inter-dependence of language and problem solving skills. Hermer-Vazquez et al. (1999), for example, have reported that children's ability to correctly solve spatial problem-solving tasks is positively associated with language competence. Leroy et al. (2012) studied the nonverbal reasoning ability of children with DLD and typically developing children, and, more specifically, examined children's solving of reasoning problems in which searching for relational similarity and perceptual similarity was important. These researchers found that the reasoning performance of children with DLD was less developed in comparison to that of typically developing children. These results suggest that language might play an important role in various forms of reasoning and solving complex problems.

Dynamic testing

Dynamic testing approaches may have value for children with learning difficulties, or who received less favourable learning opportunities in the past. Only a few studies in dynamic testing have been conducted in the field of language disorders (e.g., Camilleri & Botting, 2013; Hasson & Botting, 2010; Poehner & Lantolf, 2005, 2013). Dynamic testing, in principle, allows for observing the nature and breadth of a child's progression in performance when providing them with feedback (Hessels, 2000; Sternberg & Grigorenko, 2004). However, educational or school psychologists often do not have time to undertake highly time-consuming assessment of children, a reason why a speedy, structured approach to dynamic testing would be welcome and necessary (e.g., Resing, Touw et al., 2017). The approach utilised in the current study entails a computerised pre-test-training-post-test format with guided instruction. The training procedure employs the use of graduated prompts provided to the child only if difficulties in task solving arise (Campione & Brown, 1987; Fabio, 2005; Resing, 2000; Resing & Elliott, 2011). The prompts, provided digitally by a tablet, are restricted in the sense that the child is provided with only the minimum number of these scaffolds necessary to accurately solve the trained task. Changes in the number and quality of prompts needed during training are considered indicative of a child's potential for learning (Resing, 2013).

The graduated prompts approach utilised in the current study employed inductive reasoning tasks. These tasks, including analogical reasoning, categorisation and seriation, all require rule finding processes by searching for underlying similarities and differences between task features, or in the relations between these task aspects (e.g., Klauer & Phye, 2008). Changes in the problem solving behaviour of children undertaking inductive reasoning tasks after training have been reported in studies regarding class-inclusion tasks (Siegler & Svetina, 2006), and matrices/ analogies (Alexander et al., 1987; Siegler & Svetina, 2002; Tunteler et al., 2008). This study employed visual-spatial series completion. Dynamic testing research using series completion tasks is sparse (e.g., Ferrara et al., 1986; Resing & Elliott, 2011). To solve the incomplete series, children have to construct the next figure in a line of figures, by determining the number of changes in relationships (transformations) and the periodicity of regularities in the row.

Computerised dynamic testing

Dynamic testing has clear advantages to traditional static testing. This form of testing, for instance, reveals information about the underlying processes during learning (Elliott et al., 2010), and individual instructional needs (Resing, 2013). It has been argued that this information provides useful information on which didactic interventions can be based (Jeltova et al., 2007). Many dynamic tests follow a test-training-test design, also known as the ‘sandwich’ approach (Sternberg & Grigorenko, 2002). Such procedures, however, cost time and, therefore, are still difficult to apply in practice. Computerised dynamic testing, therefore, is gaining more attention since scores can be registered automatically, and assessment outcomes can be inspected immediately after testing. Several researchers have explored computer-assisted feedback during a dynamic test (e.g., Passig et al., 2016; Poehner & Lantolf, 2013; Resing et al., 2011; Stevenson et al., 2011; Tzuriel, & Shamir, 2002). In the current study, we utilised a newly developed dynamic series completion test using a whole-figure geometric solution instead of a multiple-parts item solution to be administered on a tablet. It was assumed that such a test on a tablet could be easily incorporated into the classroom, and could be managed with ease by young school children. In addition, children’s motivation has been found to be positively affected by using a tablet (e.g., Furio et al., 2015). Therefore, we considered a tablet to potentially be an authentic assessment instrument for young school children. In contrast to most dynamic tests, we kept verbal instructions as simple and short as possible, visually supported by the information on the tablet.

Aims of the current study

The present study aimed to investigate the usefulness of a computerised dynamic test, employing graduated prompts, as a means to provide insight into the cognitive abilities of children with language difficulties. We sought to investigate whether dynamically versus statically tested children with a developmental language disorder (DLD) showed differential patterns in the way they solved serial reasoning tasks before and after training or unguided practice. Our aim was to substantiate these earlier findings in dynamic testing and extend the results to a group of children with DLD who attended special education in the Netherlands.

Firstly, we examined the effect of graduated prompts training on children’s solving of series completion task-items, looking at the number of correctly applied transformations they utilised in their answer. Based on studies with typically developing children, we expected that trained children with DLD would improve their number of correctly applied transformations while solving reasoning tasks more than control-group children with DLD (e.g., Resing & Elliott, 2011; Resing et al., 2012), whereby we controlled for age, initial reasoning level, and ethnicity.

In addition, we expected that the weakest scoring children on the pre-test, compared to their higher scoring peers, would improve their serial reasoning skills significantly more after training. In previous studies, amongst typically developing children, it was found that in general children with initial lower scores on a dynamic test of inductive reasoning improved most (Resing, Bakker et al., 2017; Stevenson et al., 2013). We sought to study if we could substantiate these findings to children with a DLD.

Thirdly, we examined children's instructional needs patterns by examining potential individual differences in their need for prompts during training. In accordance with previous findings amongst typically developing children (e.g., Resing & Elliott, 2011), we expected that children would need fewer prompts in the second training session than in the first. We further examined the usefulness of a computerised dynamic test in revealing individual patterns of instructional needs.

4.2 Method

Participants

A total of 47 children (33 boys and 14 girls), with a mean age of 8.06 years (age range 7.17 – 9.33 years) and a standard deviation of 6.9 months, participated in this study. The children's primary language spoken at school was Dutch. All children were in the second grade of seven special education schools for children with a Developmental Language Disorder (DLD). All schools were located in the western part of the Netherlands. The schools were selected on the basis of their willingness to participate in this study. The sample was heterogeneous in terms of language difficulties, including children with a range of receptive and expressive language difficulties. Nine children were known to be diagnosed with other comorbidities (e.g., autism spectrum disorder, attention deficit disorder). For 13 children no information about comorbidities was available. The presence of comorbidities was not used as an exclusion criterion. Hearing problems were used as an exclusion criterion as these children could experience difficulty with the verbal instructions and prompts. No information about socio-economic status was collected. Written parental informed consent was obtained for the participation of all children. Initially, parental consent of 51 children was collected, but four children were absent during the administration of the post-test and were therefore excluded from the analyses.

Design

This study employed a pre-test-post-test control-group randomised blocking design. The children were randomly assigned to an experimental or a control condition by using blocking based on inductive reasoning scores from the Raven Standards Progressive Matrices (SPM; Raven, 1981), for each school separately. They attended five sessions in total. In the first session all children were administered the Raven test. Based on the score that the children obtained in the Raven test, an estimate could be made of their level of inductive reasoning. In the second session all participants received the pre-test of the dynamic test of series completion, to establish a baseline level. In the third and fourth sessions participants in the experimental condition (*n* = 23) were given training using a series completion task similar to the pre-test with graduated prompting. The participants in the control condition (*n* = 24) were given visual spatial puzzles. In the last session all participants received a post-test to measure possible improvement. During the pre and post-tests no graduated prompts were provided. The sessions were separated by about seven days. This study combined data sets of two smaller studies with DLD children attending special education. The studies used an identical dynamic test with series completion items but with a minor change in the task design. We checked whether it was possible to combine both datasets by obtaining a rank-order correlation utilising Spearman's rho between the item-sets of the pre-tests of the two studies, which appeared to be satisfactory (r_s = .90, p < .001).

Materials

Raven's Progressive Matrices. A paper-based version of the Standard Progressive Matrices (SPM; Raven, 1981), was used in this study, measuring children's fluid intelligence, specifically their inductive reasoning. The Raven test consists of 60 items, with an increasing difficulty level, which are composed of black and white figures with a missing piece, where the children had to choose the missing element of the figure. The task was administered in a group of three to four children. The Raven test has a reliability of α = .83 and a split-half coefficient of r = .91 (Raven, 1981).

Series completion test: Construction. A computerised geometric series completion test was used to measure children's seriation skills. Children were asked to complete sequential patterns consisting of a row of six geometric figures, administered on a tablet. The seventh place in the row was left empty and the children had to drag the fitting geometric figure(s) to the empty place so the series would be completed. Children had to detect the changes (transformations and periodicity) in the pattern of the row and determine the underlying rule. Five possible transformations could occur within a row: changes in geometrical shape, colour, size, quantity and position. Figure 1 shows an example item of the series completion test. Items were constructed on the basis of former findings regarding the solving of series (Holzman et al., 1983), with increasing numbers of transformations in items and the frequency of recurring periods of change (periodicity).

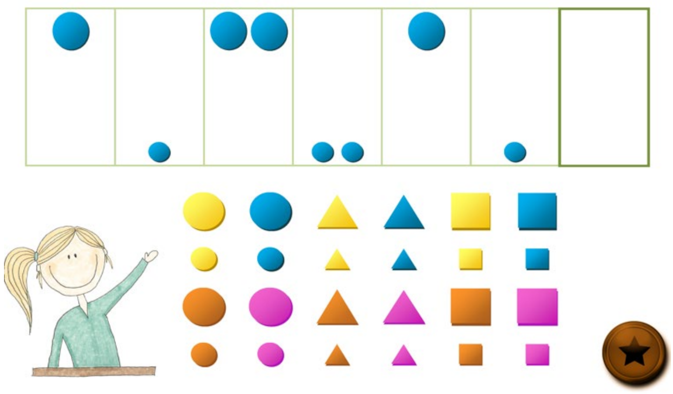


Figure 1. Geometric series completion test. A relatively easy item with 3 transformations; size (periodicity 2), quantity (periodicity 2) and position (periodicity 2).

Series completion test: Pre- and post-test. The pre- and post-test consisted of eighteen series completion items, with two additional example items at the start of the session. Before the start of the pre-test, we checked whether the children were familiar with the geometrical shapes used in the task. The pre- and post-test consisted of two parallel item sets, with identical patterns of item difficulty, but different geometric shapes (circles, triangles and squares), formats (large and small), and colours (blue, orange, pink and yellow) were used to construct an item. In order to analyse (changes in) children's solving process in detail, answers to pre-test and post-test items were scored in terms of the number of transformations children applied correctly. Each seriation item contained a number of changes in dimensions (transformations), ranging from one to five for each item, and the different transformations that children were able to apply correctly were scored and summed. This scoring system implied that even though their overall answer may not have been correct, some of the task elements were, indeed, applied correctly and were, in this way, given credit. Administration time of the pre- and post-test was about 30 to 45 minutes.

Series completion test: Training. Children in the training condition received two training sessions (session 3 and 4) consisting of six series completion items each that were constructed equivalently. If children were unable to solve the task independently, they received feedback using a standardised graduated prompts procedure. Prompts were provided hierarchically, becoming increasingly more specific, so that children were facilitated increasingly directively towards the correct answer (e.g., Campione & Brown, 1987; Ferrara et al., 1986; Resing, 2000; Resing & Elliott, 2011). The children received a maximum of four verbal prompts for each item. The first two prompts given were general metacognitive prompts, reminding the children to look at the figures and determine the changes and similarities in the row. The next prompt was a more explicit cognitive prompt, emphasising the different transformations. The final prompt consisted of direct guidance by scaffolding. After a correct answer, no further feedback was given and the tablet proceeded to the next item. The administration time was about 30 minutes per training session.

Series completion test: Visual aspects. The test was administered on an Acer Aspire Switch 10 convertible tablet, operating on Windows, with a 10.1-inch touchscreen display with a resolution of 1280 x 800 pixels. The pre- and post-test were designed such that they were highly visual. During the test sessions, the tablet provided different kinds of output: visual, verbal and sound feedback. An animated figure, named Lisa, appeared on the left side of the screen and gave the children verbal instructions. In addition, the tablet provided visual effects parallel to the verbal instructions in all four sessions to visually attract attention to the figures. The tablet briefly enlarged the geometric figures in the series, the outlines of the boxes, and the outline of the entire row. The prompts given during the training were predominantly verbal. To adapt the training to the specific group of children examined, children with DLD, the second prompt provided included visual support. Furthermore, the tablet provided auditory feedback during the example and training items. A high ‘pling’ sound was played whenever an answer was correct and a lower frequency in case of incorrect answers. In Figure 2, a schematic and detailed overview of the computerised series-completion test is presented.

Pre-test and Post-test	**Graduated Prompts Training**
Display: The tablet display visually presents the task to the child. The row with geometric figures is shown on the display of the tablet. The child can tap on a basket to reveal the geometric shapes in four different colours and two sizes (see Figure 1), select the shape he or she wants to use, and drag it to the empty box in the row of figures. When the child drags a shape into the last (empty) box of the figure he or she needs to press a star-button in order to confirm the answer and proceed to the next item (see Figure 1). Verbal instruction: An animated figure provides general verbal instructions. After each item, the children are asked why they chose their answer. Sound: The tablet provides additional auditory feedback after an answer is given during the example items of the pre- and post-test and the training procedure. A high ‘pling’ sound is played whenever an answer is correct and a lower sound when the child’s answer is incorrect. Visuals: The tablet provides visual effects parallel to the verbal instructions to visually attract attention to the figures. The tablet briefly enlarges the geometric figures in the row, the outlines of the boxes and the outline of the complete row.	
	Prompts: After each correct answer, the child receives positive feedback and is asked why they chose their answer. When an answer is incorrect, prompts are provided by the tablet.
	Prompt 1 (metacognitive): *Look at the row again. What do you have to do to complete the row?*
	Prompt 2 (metacognitive): *Look at what changes in the row and what does not. Pay attention to shape, colour, small or big, one or two, and where in the figure.* 16 of the 23 children received prompt 2 also visually: 
	Prompt 3 (cognitive, item-specific): The tablet points out the changing transformations (shape, colour, size, quantity and position) in the row, and the child is asked to try again.
	Prompt 4 (cognitive, item-specific): The tablet only points out the elements that are incorrect. If the child’s answer is incorrect again, the correct answer is shown by the tablet.

Figure 2. Overview of the different kind of output presented by the tablet during administration of the computerised series-completion test.

4.3 Results

Prior to examining our research questions, we checked whether children in the two conditions did not differ significantly in their initial level of inductive reasoning, using Raven's Progressive Matrices Test scores, age at the start of testing, and ethnicity. Three separate ANOVAs revealed that children in both conditions did not significantly differ with regard to their initial level of inductive reasoning ($F(1, 45) = .126$, $p = .724$, $\eta_p^2 = .003$), age ($F(1, 45) = .427$, $p = .517$, $\eta_p^2 = .009$), or ethnicity ($F(1, 45) = .777$, $p = .383$, $\eta_p^2 = .017$). In Table 1, an overview of the basic statistics of the two experimental groups can be found.

Table 1. Basic statistics of the children in the two conditions (Control and Training).

			N	Mean	SD
Age in months	Control		24	97.38	6.83
	Training		23	96.04	7.14
Raven raw scores	Control		24	28.71	9.63
	Training		23	27.83	7.17
Transformations correct	Control	Pre-test	24	77.92	7.62
		Post-test	24	76.67	9.32
	Training	Pre-test	23	72.87	8.83
		Post-test	23	78.43	7.69
Prompts needed		Training 1	23	16.70	4.12
		Training 2	23	14.09	5.72

Effect of the dynamic series completion test

We expected that trained children with DLD would improve their seriation skills more than control-group children with DLD. We examined the effect of training on the number of correct transformations at pre-test and post-test by conducting a repeated measure ANOVA with Condition (training/control) as between-subjects factor and Session (pre-test/post-test) as within-subjects factor. The analysis revealed an effect for Session bordering on significance (Wilks' $\lambda = .932$, $F(1, 45) = 3.3$, $p = .076$, $\eta_p^2 = .068$), and, more importantly, a significant interaction effect of Session x Condition (Wilks' $\lambda = .845$, $F(1, 45) = 8.23$, $p = .006$, $\eta_p^2 = .155$). A visual examination of the mean scores, as provided in Figure 3, indicated that the control-group children did not show any improvement in the mean number of correct transformations from pre-test to post-test, whereas trained children showed a significant increase in the mean number of correct transformations.

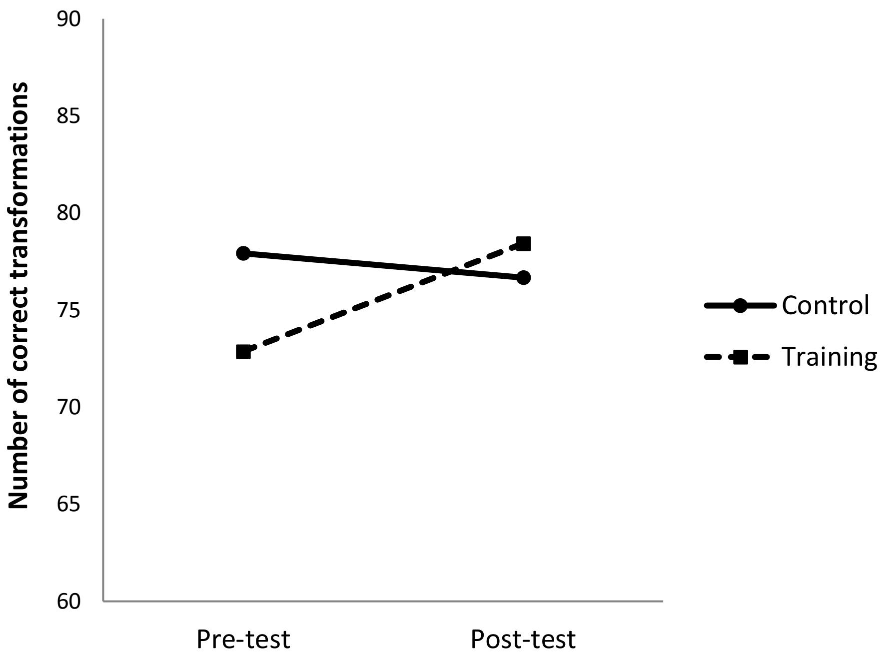


Figure 3. Mean scores of the number of correctly applied transformations during the pre- and post-test, displayed by condition.

Furthermore, we analysed whether children with weaker scores on the series completion task would benefit more from a dynamic training and, therefore, show more progression on the post-test, compared to their non-trained peers. We used the median of the total number of correctly solved items on the pre-test to create a weak performing group and a strong performing group. Five children scored at the median and were therefore excluded from the analysis. An additional repeated measure ANOVA was conducted with Condition (training/control) and Initial Performance (weak/strong) as between-subjects' factors and Session (pre-test/post-test) as within-subjects factor. A significant Session (Wilks' λ = .80, $F(1, 38) = 9.83$, $p = .003$, $\eta_p^2 = .21$) and Session x Condition (Wilks' λ = .69, $F(1, 38) = 17.73$, $p < .001$, $\eta_p^2 = .32$) effect were revealed. However, in contrast with our hypothesis non-significant Session x Initial Performance (Wilks' λ = .97, $F(1, 38) = 1.31$, $p = .260$, $\eta_p^2 = .03$), and Session x Condition x Initial Performance (Wilks' λ = .98, $F(1, 38) = .67$, $p = .420$, $\eta_p^2 = .02$) effects were found. These results indicated that there was no statistical significant difference in progression between the children with weaker and stronger initial performance, possibly due to the small number of participants. However, visual examination of the mean scores of the different subgroups, in combination with a significant between-subjects effect of Initial Performance ($F(1, 38) = 9.07$, $p = .005$, $\eta_p^2 = .19$), and a between-subjects interaction effect of Condition x Initial Performance ($F(1, 38) = 3.17$, $p = .083$, $\eta_p^2 = .08$) that is bordering on significance, portrayed in Figure 4 suggested that trained children with weaker scores on the series completion task demonstrated greater progression in series completion (correctly applied transformations) compared to the trained children with stronger initial performances. It seemed that the children scoring relatively low on the pre-test are catching up with the children scoring relatively high on the pre-test, as a result of training. Of course, the children with lower pre-test scores had more room for improvement than those with higher initial scores.

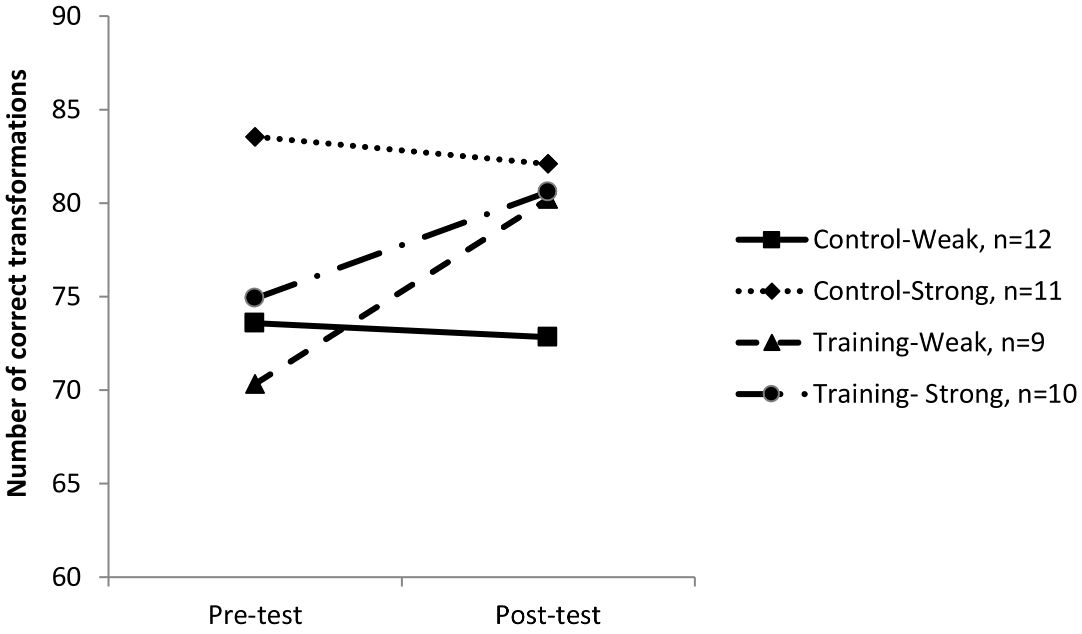


Figure 4. Mean score of the number of correctly applied transformations during the pre- and post-test, by condition and initial performance.

Individual instructional needs patterns

Next, we wanted to know more about the instructional need patterns of the dynamically trained children with DLD, as insight into these patterns may provide teachers, educational psychologists or other educational professionals working with children with DLD with useful tools for guiding these children and adapting to their cognitive skills. Considering both training sessions together, out of 48 possible prompts, the children needed an average of 30.8 (*sd* 8.7) prompts. On average they needed 16.7 (*sd* 4.1) prompts in the first training session, and 14.1 (*sd* 5.7) prompts in the second training session. A repeated measure ANOVA with Session (pre-test/post-test) as within-subjects factor, showed a significant decrease in the total number of prompts from training session 1 to 2 (Wilks' λ = .766, $F(1, 22) = 6.73$, $p = .017$, $\eta_p^2 = .234$). The total number of prompts that the children with DLD needed per training session are provided in Figure 5. When exploring the individual differences in needed prompts, large differences were found in instructional needs patterns of the children, as displayed in Figure 6. For example, some children showed a relatively large decrease in their need for prompts after the first training session, such as participant 7, which is especially salient when examining the fourth prompt, while other children showed they more or less needed the same number of prompts in both training sessions, such as participant 2.

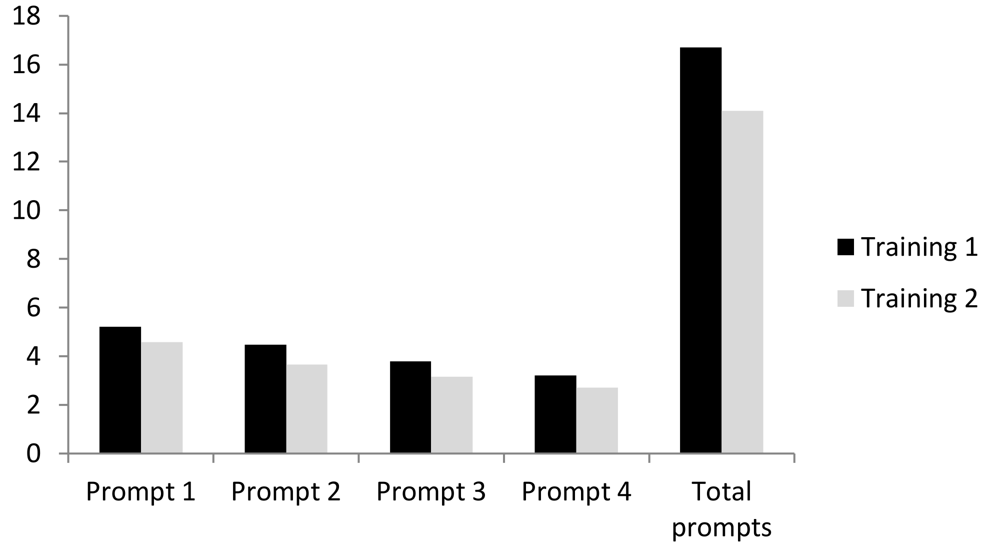


Figure 5. Mean number of prompts, divided by training session and type of prompt provided.

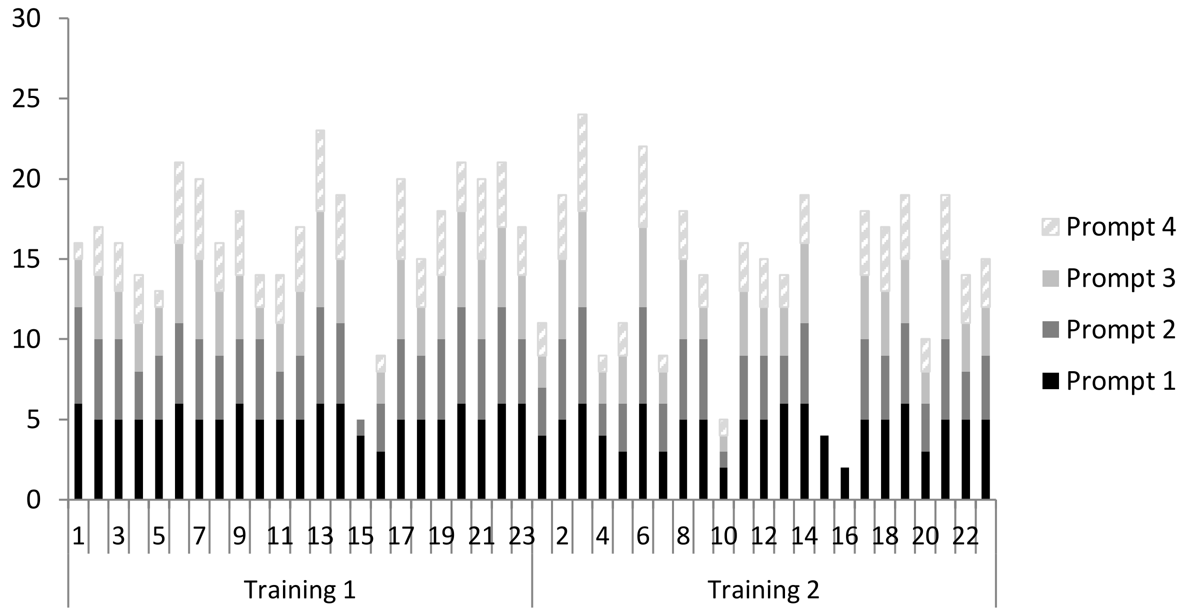


Figure 6. Individual differences in the number of prompts needed during training session 1 and 2 per child, divided in the four prompts.

4.4 Discussion

The main aim of this study was to examine the potential value of computerised dynamic testing for assessing series completion performance of children with Developmental Language Disorder (DLD).

As research in the field of dynamic testing for this population is scarce, we sought to examine whether the results of previous studies amongst typically developing children could be extended to the population of children with DLD.

Our first task was to examine whether a graduated prompts approach would lead to children displaying larger progression paths than if they were subjected to repeated practice opportunities only. Upon examining children's changes in series-completion performance, we found that the graduated prompts training brought about larger changes in accurately applied transformations than repeated practice only, as was also found in a myriad of previous studies with typically developed children (e.g., Resing & Elliott, 2011; Resing et al., 2012). These findings, therefore, provide a first indication of the usefulness of dynamic testing in unveiling the potential for learning in the population of children with DLD.

Secondly, we sought to investigate children's progression paths in more detail by dividing children into a group of weak and a group of strong initial performers, as Siegler (2007) postulated that initial task performance is a key determinant for examining cognitive change. Although a visual check of children's progression paths seemed to indicate that the children with weaker initial scores did indeed show larger progression paths after training than those with stronger initial scores, no significant differences were found in the extent to which they progressed from pre-test to post-test. A possible reason as to why we did not find any significant differences could be the relatively small number of participants in the current study. In future research, therefore, we would advise using larger sample sizes. Not only will this enhance the generalisability of the findings of the present study, it will also, hopefully, lead to more power.

Finally, we examined in detail the instructional needs patterns of the children with DLD. The current study also reveals that within the population of children with DLD, large differences exist in the instructions these children need when being trained dynamically. Mean scores further revealed, similar to previous findings with typically developing children (e.g., Resing & Elliott, 2011), that in general children needed fewer prompts in the second training session than in the first, although there were also large individual variations in the level of decrease found. These findings provide a further indication of the value of dynamic testing for the DLD population, and remind us that, just like typically developing children, children with DLD show differences in the amount and type of help they need in order to learn within their individual zone of proximal development (e.g., Resing, 2013). Dynamic testing seems to be useful for this special population in obtaining in-depth information about these individuals' needs for instruction. Teachers can, for instance, be provided with the individual patterns of instructions the children demonstrated. Some children, for example, highly profited from general, metacognitive help, whereas others have shown to profit more from more specific step-by-step scaffolds. Having this specific information might make it easier for teachers to differentiate in the type of instructions they provide to children with language difficulties. In combination with information about children's task approach, information about instructional needs can be used as an indication of these children's further educational needs, and can serve as a guideline for devising and implementing didactic interventions (Jeltova et al., 2007).

Of course, it must be noted that the training procedure utilised in the current study relied on verbal language. Although we aimed to create a procedure that emphasised several visual effects, language was used to convey the content of the prompts. Previous studies suggest that this might lead to difficulties when the child assessed has problems in the domain of language (e.g., Camilleri & Botting, 2013). Moreover, other studies revealed that language plays a key role in reasoning and problem-solving (Baldo et al., 2015; Berk & Garvin, 1984; Hermer-Vazquez et al., 1999; Lidstone et al., 2012; Winsler et al., 1997). These findings seem especially salient when taking into account Leroy et al.'s (2012) study, in which it was suggested that children with DLD seem to have difficulties in solving cognitive tasks irrespective of their individual language abilities.

We utilised a newly developed computerised dynamic test of series-completion. In educational practice, computerised assessment seems to be used increasingly (Clements & Samara, 2002). Therefore, the current study yields important implications for the field of education, and assessment in general. Not only has computerised testing been found to have several advantages for assessing children's cognitive abilities, (e.g., Tzuriel & Shamir, 2002; Williams & McCord, 2006), this form of testing also enables obtaining in-depth information about children's problem-solving behaviour during the test. Therefore, computerised testing, such as the dynamic test employed in the current study, can potentially provide a wealth of fine-grained information about children's problem-solving capacities, and instructional needs. This information, in turn, could be used by psychologists, educational professionals and researchers alike to make adaptations in educational provisions for all children, in general, and, more specifically, special populations such as children with DLD, for example in the case of issues arising in the implementation of inclusive education.

In conclusion, the current study serves as a first step in uncovering the usefulness of dynamic testing for children with DLD. A myriad of studies has shown that dynamic testing is a powerful tool in obtaining information about children's potential for learning (e.g., Resing, 2013; Resing & Elliott, 2011). As far as the authors are aware, the current study was one of the first to investigate the value of a computerised dynamic test for assessing the inductive reasoning abilities of children with DLD. Our findings indicate that this form of testing can, indeed, be used to obtain information about the potential and learning processes of children whose abilities on static assessment tools might be underestimated.