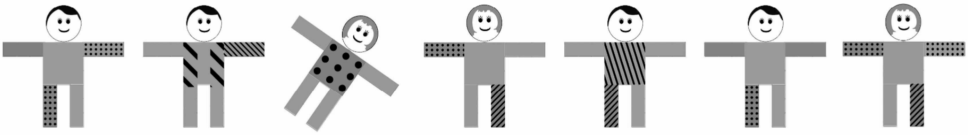
Dynamic testing and cognitive flexibility
Stad, F.E.

Chapter 4



Predicting school achievement: differential effects of dynamic testing measures and cognitive flexibility for math performance

Femke E. Stad

Carline J.M. van Heijningen

Karl H. Wiedl

Wilma C.M. Resing

Stad, F. E., Van Heijnigen, C. J. M., Wiedl, K. H., & Resing, W. C. M. (submitted).

Predicting school achievement: differential effects of dynamic testing measures and cognitive flexibility for math performance. *[Learning and Individual Differences]*

Abstract

Inductive reasoning and cognitive flexibility are both argued to be related to children's math achievements at school. The current study aimed to investigate whether dynamic measures of inductive reasoning would provide additional predictive value of math achievement while taking into account static inductive reasoning performance and cognitive flexibility. Six and seven year old children were administered a dynamic test of series completion comprising a pre-test – training – post-test format and a test of cognitive flexibility. Half of the children were trained in series completion while the other half of the children only practiced. Cognitive flexibility and the dynamic measures of inductive reasoning were each found to provide additional predictive value to static pre-test performance and to hold unique predictive value for math achievement. The results underline the importance of both dynamic testing and cognitive flexibility in educational assessment.

For teachers and other education professionals it has become increasingly important to monitor the child's learning progression and individual needs in order to provide an optimal learning environment for each child (e.g., Pameijer, 2017; Stecker, Fuchs, & Fuchs, 2008; Resing, 2013). Individual or group-wise assessment of cognitive development has become increasingly important and often includes traditional measurements of cognitive functioning that focus on measuring previously acquired knowledge and skills at a particular moment in time (e.g., Sternberg & Grigorenko, 2002; Resing, 2000). Although these traditional measurements are certainly valuable for their predictive qualities regarding school performance (e.g., Lewis, 2013; Neisser et al., 1996; Roth et al., 2015; Sternberg, Grigorenko, & Bundy, 2001), these test outcomes might not suffice in providing information regarding the child's potential for learning and instructional needs in order to further unfold this potential (e.g., Fuchs, Compton, Fuchs, Bouten, & Caffrey, 2011; Haywood & Tzuriel, 2002; Jeltova et al., 2011). An additional way of cognitive measurement that aims to explore the child's ability to profit from feedback and instruction is dynamic testing, a form of testing that includes a training situation in the assessment process in order to not only examine the level, but also the rate and process of learning (e.g., Carlson & Wiedl, 1992; Sternberg & Grigorenko, 2002). Different forms of dynamic testing can be distinguished based on the way feedback and help are provided during the intervention phase. Depending on the format and the structure of the test, dynamic test data can provide us with an understanding of the child's need for and ability to profit from feedback and instruction, providing an opportunity to gain an understanding of the child's potential for learning (e.g., Elliot, Grigorenko, & Resing, 2010).

One particular approach of providing structured feedback during the testing process is called the 'graduated prompts technique' (e.g., Campione & Brown, 1987). The graduated prompts procedure has been developed to offer a framework for helping individuals to solve test items independently by providing prompts during the training phase that gradually change from broad, metacognitive hints to more detailed, step-by-step hints, based on the individual's need for help (e.g., Resing & Elliott, 2011). This approach is often embedded in a pre-test – training – post-test design, a dynamic test format that allows the examiner to explore the child's progression after training and the amount of feedback required to achieve this progress (e.g., Elliott, 2003; Resing, Elliott, & Grigorenko, 2012).

The purpose of the current study was to investigate whether dynamic measures, derived from the graduated prompts technique, were unique predictors of children's school achievement, as compared to the predictive value of static pre-test performance.

Dynamic tests often incorporate inductive reasoning measures, in which knowledge about a particular situation is used to infer generalizations regarding new situations (e.g., Ferrara, Brown, & Campione, 1986; Resing, 2000). Inductive reasoning can be considered to be a core element in much of learning in school (e.g., Goswami, 1996; Esposito & Bauer, 2017; Goswami, 2013; Perret, 2015). The ability to solve inductive reasoning tasks, often considered a measure of fluid reasoning, has been shown to be a good predictor of school achievement in both reading and math (e.g., Ferrer et al., 2007; Taub, Keith, Floyd, & McGrew, 2008). Dynamic measures of inductive reasoning have been found to provide additional information on children's present and future attainment at school, as compared to performance measures on standardized achievement tests (e.g., Caffrey, Fuchs, & Fuchs, 2008; Hamers, Pennings, & Guthke, 1994; Resing, 1993; Spector, 1992; Swanson, 1994; Tissink, Hamers, & Van Luit, 1993). In addition, the predictive validity of post-training scores on school grades appears to be fairly high; post-test scores tend to show significantly higher correlations with school performance than do static pre-test scores (e.g., Budoff, 1987; Guthke & Wingenfeld, 1992). This predictive validity of dynamic testing on standardized school achievement scores is important because it contributes to identifying students at risk for school failure and in need of more intensive intervention (e.g., Caffrey et al., 2008).

However, not only static and dynamic measures of inductive reasoning have been found to hold predictive value for children's current and future school achievement. Children's executive functions, an umbrella term for the cognitive processes that are responsible for purposeful and goal-directed behaviour such as planning, reasoning, and monitoring (e.g., Anderson, 2001; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003), have been shown to be associated with children's performance on standardized measures of school-related subjects. Children with better executive functioning performed better on static measures of math and reading (e.g., Gathercole, Pickering, Knight, & Stegmann, 2004; Roebers, Cimeli, Röthlisberger, & Neuenschwander, 2012; St Clair-Thompson & Gathercole, 2006) as compared to their peers with weaker executive function skills. This is not surprising as executive functions have been found to be fairly strongly related to inductive reasoning abilities (e.g. Cho, Holyoak, & Cannon, 2007; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Executive functions are often divided into working memory, inhibition and cognitive flexibility (e.g., Anderson, 2002; Diamond, 2013; Miyake et al., 2000). Where previous research has mainly highlighted the predictive relationship between working memory and school achievement (e.g., Alloway & Alloway, 2010; Alloway & Passolunghi, 2011; Krumm, Ziegler, & Buehner, 2008), the role of inhibition control and cognitive flexibility has not been studied so detailed. The current study focused on examining whether cognitive flexibility held unique predictive value for children's school achievement in math, while taking into account predictive values of static and dynamic measures of inductive reasoning. Cognitive flexibility is characterized by the ability to alter or modify previous learned behaviours and to learn from mistakes, whereas inflexibility is characterized by both perseverative behaviour and the inability to adapt to new task demands or utilize feedback (e.g., Anderson, 2002; Diamond, 2013). Cognitive flexibility has been argued to build on the other two functions and shows significant developmental changes around the age of 6-7 (e.g., Davidson, Amso, Anderson, & Diamond, 2006; Gupta, Kar, & Srinivasan, 2009; Heaton, Chelune, Talley, Kay, & Curtiss, 1993; Huizinga, Burack, & Van der Molen, 2010). Changes in cognitive flexibility coincide with the first two years of formal education and might therefore be of particular interest when exploring children's potential for learning (e.g., Stevenson, Bergwerff, Heiser, & Resing, 2014). Cognitive flexibility (or shifting ability) has been argued to be particularly related to performance in math-subjects because these subjects explicitly require switching between different aspects of the task or arithmetical strategies (e.g., Agostino, Johnson, & Pascual-Leone, 2010; Blair, Knipe, & Gamson, 2008; Bull & Scerif, 2001; Clark, Pritchard, & Woodward, 2010; Yeniad, Malda, Mesman, Van IJzendoorn, & Pieper, 2013).

The current study sought to examine the predictive value of cognitive flexibility and dynamic measures of inductive reasoning derived from the graduated prompts training: progress after training and need for instruction. The aim of the study was to examine whether cognitive flexibility performance and dynamic measures of inductive reasoning could be considered unique predictors of young children's math achievement at school. Primary school children were tested with a dynamic series completion test, and their performance on national scholastic achievement tests of math was collected. Although our primary research aim was to gain insight into the (unique) predictive values of dynamic measures and cognitive flexibility, we firstly examined the effectiveness of the graduated prompts training in improving children's inductive reasoning ability. Because previous studies utilizing the dynamic series completion test have shown the effectiveness of this particular training procedure in improving children's reasoning skills, our first hypothesis was that children receiving the series completion training would show significantly more progress in their inductive reasoning ability from pre- to post-test than the children not receiving training (e.g., Resing & Elliott, 2011; Resing, Tunteler, & Elliott, 2015; Resing, Xenidou-Dervou, Steijn, & Elliott, 2012). A second research aim was to examine whether the dynamic measures derived from the dynamic series completion test would provide additional predictive value over the static performance measure. Based on existing dynamic testing literature (e.g., Caffrey et al., 2008; Haywood & Lidz, 2006; Lidz, 1991) we expected that dynamic test performance after training and number of prompts required during training would provide significant predictive value in addition to static pre-test performance. The third and last research aim concerned whether dynamic measures of series completion would provide predictive value for children's math achievement while taking into account cognitive flexibility. Although previous studies have shown that dynamic measures of inductive reasoning and cognitive flexibility performances are both substantially associated with children's math achievement (e.g., Agostino et al., 2010; Ferrer et al., 2007), to date it remains unclear whether these variables hold unique predictive value for math achievement or that the predictive value of the dynamic test measures could be largely related to children's cognitive flexibility skills. We therefore explored the extent to which dynamic measures of inductive reasoning hold unique predictive value for math achievement while taking into account the influence of cognitive flexibility.

4.2 Method

Participants

Participants were 205 native Dutch speaking children (100 boys, 105 girls) from first and second grade, who were selected from 5 middle-class primary schools in the Netherlands. The mean age of the participants was 6 years, 8 months (*SD* = 7 months). Written informed consent was obtained from the parents. Due to absence during one or more of the testing sessions, 4 children were excluded from the sample.

Design & Procedure

In the current study, a pre-test – training – post-test control-group design with randomized blocking was used. A visual exclusion test was used to block the participants based on their initial level of inductive reasoning. Based on this blocking procedure, pairs of children were randomly assigned to one of two conditions: (1) a training group in which children received training via a graduated prompts procedure and (2) a practice-control group in which children completed dot-to-dot tasks. During the first session, all children completed the visual exclusion test and a test of cognitive flexibility. During the following four sessions, the pre-test, two training sessions and the post-test of the dynamic series completion test were administered. All tasks were administered individually. Each session took place at the children's school and lasted approximately 30 minutes.

Materials

Visual Exclusion Task. The RAKIT subtest Visual Exclusion (Resing, Bleichrodt, Drenth, & Zaal, 2012) was used as a measure of children's initial inductive reasoning ability. The children were asked to detect an underlying rule in order to determine which of four abstract figures did not belong.

Modified Wisconsin Card Sorting Test. The Modified Wisconsin Card Sorting Test (Nelson, 1976; Schretlen, 2010) was used as a measure of cognitive flexibility. Children were asked to sort 48 cards (based on 4 stimulus cards) according to one of the following sorting criteria: color, shape, or number. The child received feedback on the correctness of the sort, without suggestions regarding the sorting rule. According to Nelson's (1976) criteria, the experimenter changed the sorting criterion after six consecutive correct sorts and explicitly informed the child about this switch. The first two sorting criteria were determined by the child, thereby automatically determining the third sorting criterion. The three criteria were requested in the same order during the remaining part of the test. The procedure was completed after the child sorted each criterion correctly twice or after sorting all 48 cards. Cognitive flexibility was operationalized as the percentage of perseverative errors, with a higher percentage representing less cognitive flexibility. A perseverative error was made (1) when a child incorrectly used the previous correct sorting criterion or (2) when a child did not switch between sorting rules after being informed that the sorting criterion had changed (Cianchetti, Corona, Foscoliano, Contu, & Sannio-Fancello, 2007).

Math achievement. Teachers were asked to report on children's school performance. They provided information about mathematics from the national scholastic achievement assessments (CITO; Hollenberg, Van der Hubbe, & Sanders, 2011). The CITO scores are used to track children's performance on school subjects across grades in primary school. School performance was measured on a five-point Likert scale (*1 = E, 2 = D, 3 = C, 4 = B, 5 = A*), based on national norms per age-group. The five categories are used to classify children's level of school performance (Hollenberg et al., 2011). An 'A'-score indicates a level of performance within the highest 25%. A 'B'-score indicates a performance above the average level of performance (between the 26th and 50th percentile). A 'C'-score indicates a performance below the average level of performance (between the 51st and 75th percentile). A 'D'- or 'E'-score indicates a level of performance within the lowest 25% (respectively between the 11th and 25th percentile and the lowest 10%). The reliability coefficients of the national scholastic assessments (CITO) for math in first and second grade were good (>.91) (Janssen, Verhelst, Engelen, & Scheltens, 2010).

Series completion task. A dynamic series completion task was used as a measure of children's inductive reasoning skills. The design of this task was based on the construction principles and analytic model that have previously been described in Resing and Elliott (2011) and in Resing, Touw, Veerbeek, and Elliott (2017). The task in the current study used similar guidelines and graduated prompts procedures. A series of schematic puppets was provided in each item. The schematic puppets were composed of different elements: gender (male, female), color of body parts (blue, green, yellow, and pink), and design of body parts (stripes, dots, none). These different elements changed within and across series. The children were required to encode the different elements and identify the changing relationship between the task elements in order to complete the schematic puppets series. An item example is depicted in Figure 1. They were asked to construct their solution of the last puppet in each series on a plasticized paper puppet, using eight plastic body parts. No feedback regarding the correctness of the child's solutions was provided. After each solution, the child was asked to explain his/her solution.

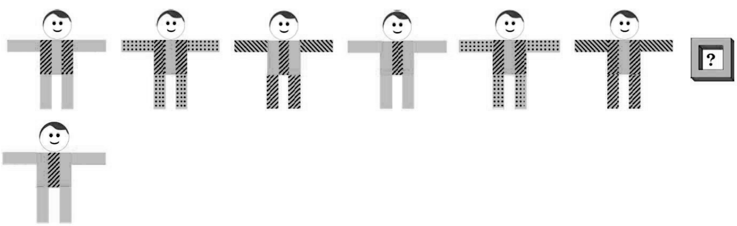


Figure 1. Item example (first row) and correct answer (the puppet below)

Series completion: pre- and post-test. The pre- and post-test each consisted of 12 series completion items, increasing in difficulty. The item difficulty was determined by two aspects: the frequency of recurring patterns (periodicity) and the number of transformations. Items identical in periodicity and number of transformations were constructed for pre- and post-test to ensure similar item difficulties across sessions.

Series completion: dynamic training. The dynamic training consisted of two training phases, in which two times six items (increasing in difficulty) were presented. Before each item, a general instruction was given. If a child encountered difficulties while constructing the correct answer, help was provided according to a standardized graduated prompts procedure. The first prompt provided after an incorrect answer was a general, metacognitive prompt. Next, if the child was not able to construct the correct answer based on the metacognitive prompt, two more specific cognitive prompts could be provided. Finally, step-by-step instructions were provided to guide the child towards a correct solution. A flowchart of the procedure is depicted in Figure 2. After each correct solution, the child was asked to explain his/her reasoning. The dynamic test was administered by undergraduate psychology students, who had received extensive training in the dynamic testing procedures.

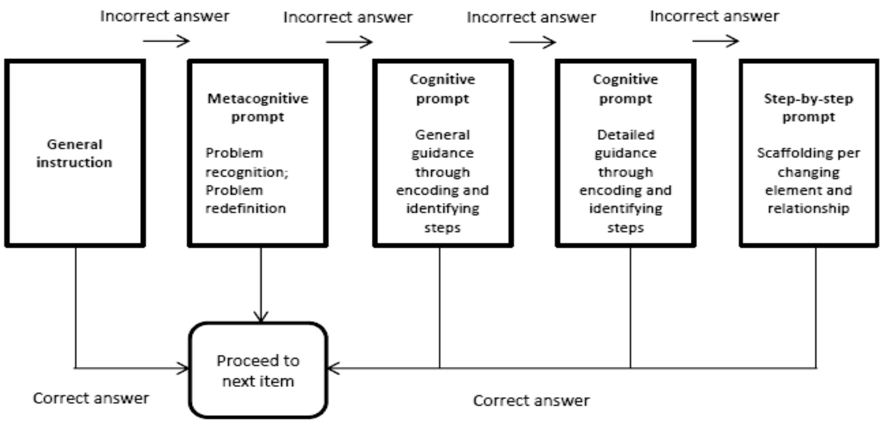


Figure 2. Flowchart of graduated prompting procedure

Analyses

Two measures of children's individual performance were derived during the dynamic series completion test: (1) Learner Status and (2) Number of Prompts required during training. The child's learner status was determined by the number of correct solutions on the pre- and post-test.

Children's gain scores, post-test minus pre-test scores, are often used to examine children's performance change after cognitive training. However, gain scores can provide an unreliable representation of children's progression from pre- to post-test as they do not sufficiently reflect the child's pre-test level and do not account for regression to the mean (e.g., Embretson & Reise, 2000; Guthke & Wiedl, 1996). Therefore, in this study, a typicality analysis was applied to the data that classified participants in 'Non Learner', 'Learner' and 'High Scorer', enabling a group-wise comparison of performance gain. This typicality analysis was applied using a pragmatic standard deviation rule of thumb (e.g., Waldorf, Wiedl, & Schöttke, 2009) that was found to be valid in classifying participants according to their learner status (e.g., Waldorf et al., 2009; Wiedl, Wienöbst, Schöttke, Green, & Nuechterlein, 2001). *Learners* were those subjects who improved their performance from pre-test to post-test by 1.5 *SD*. *High scorers* were identified as those children who scored between the pre-test upper level minus 1.5 *SD* on the pre-test. *Non Learners* did not meet either of these criteria. Multinomial logistic regression models were used because of the categorical nature of the learner status variable.

4.3 Results

Initial group comparisons and psychometric properties

Children's gender ($\chi^2(1) = .01$, $p = .93$), age ($F(1, 200) = .15$, $p = .70$), and initial inductive reasoning skills as measured by the visual exclusion task ($F(1, 200) = .25$, $p = .62$) did not differ between conditions prior to the dynamic testing process. Additionally, no differences between conditions were found regarding performance on the pre-test ($F(1, 200) = .28$, $p = .60$) and cognitive flexibility performance as measured by the M-WCST ($F(1, 200) = 1.36$, $p = .25$). Furthermore, children's scores on the national scholastic math achievement test (CITO math) did not differ between conditions ($\chi^2(4) = 3.56$, $p = .47$). Table 1 shows the descriptive statistics of pre- and post-test performance scores per condition and learner status.

Table 1. Means and standard deviations of pre-test and post-test performance on series completion per Condition (upper part Table) and Learner Status (lower part Table)

			Pre-test		Post-test	
Condition	Learner Status	N	*M*	*SD*	*M*	*SD*
Graduated prompts		125	4.53	1.85	7.43	1.92
Practice-control		76	4.16	2.01	5.16	1.88
Graduated prompts	Non Learner	43	4.96	1.48	6.04	2.11
	Learner	72	3.48	2.12	7.93	1.90
	High Scorer	10	10.3	1.49	9.90	1.20
Practice-control	Non Learner	52	3.90	1.92	4.01	1.79
	Learner	20	3.88	2.42	7.50	1.93
	High Scorer	4	9.00	1.15	8.50	2.89

For both conditions, an internal consistency (Cronbach's α) of .74 was found for the pre-test and .78 for the post-test. The correlation between the total correct scores on the pre-test and post-test for children in the practice-control condition, used as an indicator of test-retest reliability, was $r = .78$, $p < .001$.

Effectiveness of the graduated prompts training

The dynamic training was expected to improve children's performance on the series completion task. A repeated measures analysis of variance (RM ANOVA) was conducted with Performance Scores on the series completion task as dependent variable, Session (pre- and post-test) as within-subjects factor, and Condition (training and practice-control group) as between-subjects factor. The main effect of Session was significant (Wilks's $\lambda = .73$, $F(1, 200) = 53.28$, $p <.001$, $\eta_p^2 = .27$), which indicated that children showed, on average, significantly higher performance scores on the post-test than on the pre-test, regardless condition. More importantly, the interaction effect between Session and Condition was significant (Wilks's λ = .92, $F(1, 200) = 12.60$, p <.01, η_p^2 = .08), indicating that children in the graduated prompts training group, as expected, showed significantly greater progress in accuracy on the series completion task from pre- to post-test than those in the practice-control group (see Figure 3).

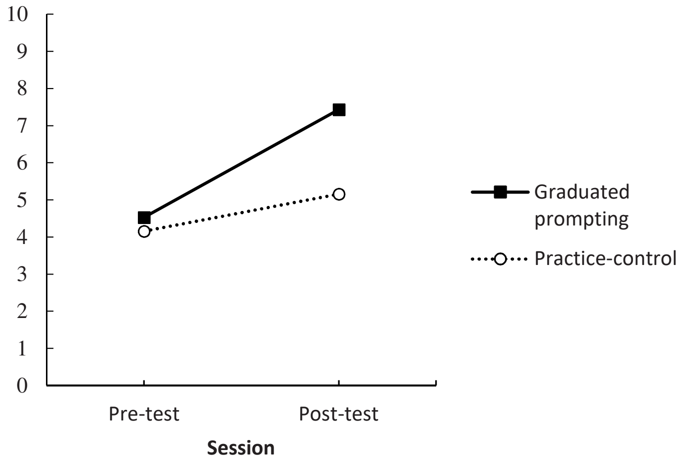


Figure 3. Patterns of change from pre- to post-test in series completion performance for trained and practice-control children

A binary logistic regression analysis was conducted to investigate whether training influenced children's learner status, with Learner Status (Learner, Non Learner and High Scorer) as dependent variable and Condition as factor. Children's classification according to Learner Status was based on a pragmatic 1.5 *SD* rule of thumb (Waldorf et al., 2009). *Leaners* were those subjects who improved their performance from pre-test to post-test with 1.5 *SD*, *High Scorers* were those subjects who scored between the pre-test upper level of 12 minus 1.5 SD, and *Non Learners* did not meet either of these criteria (see Table 1). In the current study, 92 children were classified as Learner, 95 children were classified as Non Learner, and 14 children were classified as High Scorer. The results showed that Condition significantly predicted the classification of children as Learners or Non Learners (b = 1.13, *Wald* $\chi^2(1) = 10.14$, $p < .01$). The odds ratio indicated that the chance to be classified as Learner was 3.3 times higher for a child in the training condition than for a child in the practice-control condition. Condition did not significantly predict whether children were classified as High Scorer. The results supported our hypothesis that the graduated prompts procedure improved children's series completion ability. However, this was not the case for children who already had obtained significantly higher scores on the pre-test as compared to the other children.

Predictive validity of the dynamic outcomes

To answer the question whether the dynamic measures made a significant additional contribution to the prediction of the standardized achievement scores, stepwise multinomial logistic regression analyses were carried out with Math Performance as dependent variable. In these analyses, Pre-test Scores were entered as the first and dynamic measures Learner Status and Number of Prompts as the second predictors. The Learner Status classification *High Scorers* was not included in the logistic regression analyses due to the low number of *High Scorers* per CITO category. Scores on CITO math were categorized in 4 achievement intervals: ‘A’ – the 25% highest scoring students, ‘B’ – the 25% scoring between well to just above national average, ‘C’ – the 25% scoring between just and well below national average, and ‘D’ – the 25% lowest scoring students. Basic statistics are reported in Table 2. For all analyses, achievement categories (A, B, C) were compared against the D category. The multinomial regression results are reported in Table 3.

The results for model 1 revealed that Pre-test Score and Learner Status each improved model fit and were therefore both considered predictors of math achievement. Learner Status was found to be an additional predictor of math achievement over static Pre-test Score. The results for model 2 showed that Pre-test Score and Number of Prompts both improved model fit, but that this time the dynamic measure Number of Prompts explained most of the variance in math achievement. For model 3 it was found that Pre-test Score did not add significantly to the prediction model when both Learner Status and Number of Prompts were included. Learner Status and Number of Prompts were again found to be significant predictors of math achievement, but their interaction did not improve model fit. The results indicated that, although Learner Status and Number of Prompts both significantly predicted math achievement, the success of Non Learners or Learners on the CITO math test did not depend on the number of prompts they required during training.

Although the model coefficients showed us the significant contribution of the various predictors, closer inspection of the data was needed to determine the direction of the predictor effects. Parameter coefficients are reported in Table 4. For model 1 it was found that as the variable Pre-test Score increased, the change in the odds of a child obtaining a higher score than a D score increased. It was also found that as Learner Status changed from Non Learner to Learner, the odds of a child obtaining a higher score than a D score increased. For model 2 it was found that the parameter coefficients for Pre-test Score were non-significant. For Number of Prompts it was found that when children required less hints during training, their chance of obtaining a higher score than a D score increased. Model 3 did not show a significant interaction effect in addition to the already established main effects, and parameter coefficients of the interaction effect were therefore not reported in the logistic model.

Table 2. Number of children per CITO math category, divided by Condition and Learner Status

			CITO category			
Condition	Learner Status	*N*	A	B	C	D
Graduated prompts	Non Learner	43	9	11	13	10
	Learner	72	23	22	11	16

Table 3. Overview of the forward entry logistic model coefficients examining the effects of hypothesized predictors of Math Performance (Pre-test Score, Learner Status and Number of Prompts during training)

		Model Fitting Criteria			LR test	
Model	Predictors	AIC of reduced model	BIC of reduced model	-2LL of reduced model	*df*	χ^2
Math						
1	Intercept	147.72	171.94	129.72		
	+ Pre-test Score	170.74	186.89	158.74	3	29.03***
	+ Learner Status	162.20	178.35	150.20	3	20.49***
2	Intercept	245.33	261.48			
	+ Pre-test Score	241.29	257.44	229.29	3	10.07**
	+ # Prompts	251.81	267.96	239.81	3	20.59***
3	Intercept	241.52	265.74	223.52		
	+ Pre-test Score	.	.	.	.	.
	+ Learner Status	255.62	271.77	243.62	3	20.11***
	+ # Prompts	274.69	290.84	262.69	3	39.17***
	+ Learner Status* # Prompts	.	.	.	.	.

Note. Model 1 $\chi^2(6) = 46.91$, $p < .001$; Model 2 $\chi^2(6) = 47.02$, $p < .001$; Model 3 $\chi^2(6) = 57.06$, $p < .001$.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 4. Parameter coefficients of the logistic regression models for Pre-test Score and Learner Status (model 1); and Pre-test Score and Number of Prompts during training (model 2)

				95% CI for Odds Ratio		
Model	Variable	Cito score	b (SE)	Lower	Odds Ratio	Upper
1	Pre-test Score	Category C vs D	.42 (.23)	.97	1.52	2.40
		Category B vs D	.65 (.24)**	1.20	1.91	3.03
		Category A vs D	.64 (.22)**	1.22	1.89	2.91
	Learner Status	Category C vs D	1.31 (1.23)	.33	3.70	4.89
		Category B vs D	-.72 (1.14)	.05	.49	4.51
		Category A vs D	-1.60 (1.20)***	.02	.20	2.15
2	# Prompts	Category C vs D	- .15 (.07)**	.76	.86	.98
		Category B vs D	-.17 (.06)**	.74	.84	.95
		Category A vs D	-.27 (.07)***	.66	.77	.88

Note. Model 1: R^2 = .33 (Cox & Snell), .35 (Nagelkerke). Model 2: R^2 = .30 (Cox & Snell), .32 (Nagelkerke). * p < .05, ** p < .01, *** p < .001

Predictive validity of M-WCST performance and dynamic measures

To compare the unique contributions of cognitive flexibility and dynamic test outcomes for the prediction of math performances, and to explore a possible interaction effect between the predictors, several multinomial logistic regression analyses were carried out. The dynamic measures Learner Status and Number of Prompts were entered as first, and M-WCST Performance as second predictor. Because we examined a possible interaction term, the dynamic measures and M-WCST Performance were forced into the model as main effects. Dynamic measure*M-WCST Performance was then entered as a third predictor via a stepwise procedure so that a possible interaction would only be entered into the model if it was a significant predictor of math performance. Results of the regression models are depicted in Table 5.

Table 5. Overview of the forward entry logistic model coefficients examining the effects of hypothesized predictors of Math Performance

		Model Fitting Criteria			LR test	
Model	Predictors	AIC of reduced model	BIC of reduced model	-2LL of reduced model	*df*	χ^2
Math achievement						
1	Intercept	213.93	238.16	195.93		
	+ Learner Status	227.53	243.68	215.53	3	19.60***
	+ M-WCST	247.69	263.84	235.69	3	39.76***
	+ Learner Status*M-WCST	.	.	.	.	.
2	Intercept	299.94	316.09	287.94		
	+ # Prompts	255.20	271.35	243.20	3	14.14**
	+ M-WCST	256.30	272.44	244.30	3	15.23**
	+ # Prompts*M-WCST	.	.	.	.	.
3	Intercept	226.51	274.96	190.51		
	+ Learner Status	.	.	.	.	.
	+ # Prompts	235.00	275.37	205.00	3	14.48*
	+ M-WCST	244.17	284.55	214.17	3	23.66***
	+ Learner Status* # Prompts*M-WCST	.	.	.	.	.

Note. Model 1 $\chi^2(6) = 57.64$, $p < .001$; Model 2 $\chi^2(6) = 52.18$, $p < .001$; Model 3 $\chi^2(15) = 94.32$, $p < .001$. * $p < .05$, ** $p < .01$, *** $p < .001$

The results for model 1 showed us that Learner Status and M-WCST Performance each improved model fit and were therefore considered unique predictors of math achievement. M-WCST Performance was found to explain more variance in math performance than Learner Status. No significant interaction effect was found, indicating that the success of Non Learners vs. Learners on the CITO math test did not depend on their M-WCST achievement. Similar results were found for model 2, which indicated that Number of Prompts and M-WCST Performance both uniquely improved model fit. M-WCST Performance was again found to explain more variance than the dynamic measure. Model 3 revealed that when Number of Prompts and M-WCST Performance were added to the model, Learner Status did not explain any unique variance in math performance. These results suggested that Number of Prompts and M-WCST Performance were the most important predictors for math performance. No significant interaction was found between the predictors, indicating that the success of Non Learners vs. Learners on the CITO math test did not depend on their number of prompts required during training, nor on their M-WCST achievement. Closer inspection of the data in model 3 revealed negative relations for Number of Prompts and CITO Math Performance and M-WCST Performance and CITO Math Performance, indicating that less prompts required during training (more efficient learning during training), and less perseverative errors on the M-WCST (better cognitive flexibility), predicted better math achievement. Parameter coefficients for the predictive values are reported in Table 6.

Table 6. Parameter coefficients of the logistic model for Learner Status, Number of Prompts and M-WCST Performance on Math Performance

			95% CI for Odds Ratio		
Math achievement	Variable	b (SE)	Lower	Odds Ratio	Upper
Category C vs D	# Prompts	-.22 (.09)*	.67	.81	.96
	M-WCST	.04 (.04)	.97	1.04	1.11
Category B vs D	# Prompts	-.17 (.08)*	.73	.85	.98
	M-WCST	-.04 (.03)*	.90	.96	1.02
Category A vs D	# Prompts	-.23 (.08)**	.68	.80	.94
	M-WCST	-.09 (.04)*	.85	.92	.99

Note. R^2 = .49 (Cox & Snell), .52 (Nagelkerke). * $p < .05$, ** $p < .01$, *** $p < .001$

The overall outcomes suggested that the two dynamic test parameters, Learner Status and Number of Prompts, both predicted children's math performance as measured with the CITO test, over children's static Pre-test Score. However, when M-WCST Performance was included in the analyses, M-WCST Performance and Number of Prompts required were found to be the most important predictors of math achievement. Both predictors explained unique variability in math performance, where less prompts during training and less perseverative errors on the M-WCST were related to better math performance.

4.4. Discussion

Educational assessment, both static and dynamic, aims to enable teachers and other education professionals to evaluate current school achievement and to predict future achievement (e.g., Caffrey, Fuchs, & Fuchs, 2008; Swanson & Lussier, 2001). In the current study we sought to examine the unique predictive values of cognitive flexibility and static and dynamic measures of series completion ability on young children's school achievement in math.

Part of our findings supported the results of previous studies and showed that dynamic measures of inductive reasoning, learner status and the number of prompts required during training, provided substantial predictive value of school achievement in math in addition to static pre-test performance (e.g., Beckmann, 2006; Jeltova et al., 2011; Stevenson et al., 2014). The number of prompts appeared to be the strongest predictor as compared to learner status and pre-test scores, thereby overshadowing the unique contribution of static pre-test performance. This finding was in line with previous research regarding the predictive values of dynamic measures on young children's school achievement, where it was found that children's instructional needs during training were the best predictor for reading achievement (Stevenson et al., 2014). Our results appear to underline the importance of this particular dynamic measure for the prediction of children's future school achievement.

With regard to cognitive flexibility it was concluded that better cognitive flexibility performance predicted better scores on the scholastic math test. In addition to that, and in addition to what has been reported to date, cognitive flexibility appeared to hold unique predictive value for children's math achievement while taking into account the dynamic measures of inductive reasoning. More specifically, we found that children's cognitive flexibility performance explained most of the variance in math achievement, and somewhat overshadowed the predictive value of children's learner status. These results supported the notion that the ability to efficiently switch between different task demands and problem solving strategies is an important cognitive skill in math performance, and that it can act as an unique influence on math achievement (e.g., Agostino, Johnson, & Pascual-Leone, 2010; Blair, Knipe, & Gamson, 2008). A suggestion for future studies on the predictive value of the graduated prompts training might therefore be to integrate prompts in the training procedure that aim to support possible weaknesses in cognitive flexibility. By including a more explicit assessment of the flexibility skills of a child in the testing process, the predictive value of the dynamic test outcomes might be increased.

A point for discussion concerns the ecological validity of the results. Although it appears that children's need for instruction and their level of cognitive flexibility can be used to make predictions about future math achievements, these approaches have the potential to improve. The series completion measures we used for prediction were dynamic, whereas the criterion variable math achievement was not. This restricted the ecological validity of the outcomes (e.g., Sternberg & Grigorenko, 2002; Wiedl & Herrig, 1978). Ecological validity could be improved by matching the method of assessment to the method of teaching, and it is therefore suggested that future research should focus on measuring the predictive validity of the dynamic measures with regard to single students or small groups of students whom receive adaptive and individualized teaching. The predictive power of the M-WCST could be improved in future studies by applying a dynamic version of the card sorting test. Children would then be increasingly challenged to perform in their Zone of Proximal Development, thereby improving the match between the card sorting test on the one hand and the dynamic testing assessment procedure on the other. The applicability of a dynamic version of the WCST has already been validated for clinical subjects in the field of psychoeducation and skills training (Wiedl, 1999) and for assessing learning potential in brain injury rehabilitation (Boosman et al., 2014).

Another point for discussion refers to the measurement of cognitive flexibility in the current study. Although the literature supports the assessment of flexibility through card sorting tests (e.g., Diamond, 2013), we would suggest to include other tasks aiming to measure cognitive flexibility or shifting in a follow-up study to obtain a more valid and general representation of children's level of flexibility. Cognitive flexibility has been argued to be a multi-component function (e.g., Anderson, 2002), indicating that different measures of flexibility might contribute to a different or more complex picture of children's actual level of performance.

Nevertheless, the findings of the current study suggest that cognitive flexibility and the ability to learn as measured through dynamic testing are at least to some extent separate cognitive constructs that are uniquely related to young children's math performance. These findings are twofold: on the one hand they underline the importance of developing cognitive flexibility in educational contexts. On the other hand they provide support for the usefulness of dynamic testing and of the graduated prompts procedure in particular. Not only does the graduated prompts training provide important information about the type of instructions that best aid a child's learning (e.g., Resing, 2000; Resing & Elliott, 2011; Campione & Brown, 1990), it also holds unique predictive qualities for children's future achievements. As such, the dynamic measures derived from this particular prompts procedure can contribute to a better understanding of individual needs, and thereby to creating a more optimal learning environment.