Differences in need for instruction: Dynamic testing in children with arithmetic difficulties

Bosma, T.; Stevenson, C.E.; Resing, W.C.M.

DOI
10.11114/jets.v5i6.2326
Publication date
2017
Document Version
Final published version
Published in
Journal of Education and Training Studies

Journal of Education and Training Studies
Vol. 5, No. 6; June 2017
ISSN 2324-805X E-ISSN 2324-8068
Published by Redfame Publishing
URL: http://jets.redfame.com

Differences in Need for Instruction: Dynamic Testing in Children with Arithmetic Difficulties

Tirza Bosma[1], Claire E. Stevenson[2], Wilma C. M. Resing[3]

[1]LEARN! Research Institute for Learning and Education, Vrije Universiteit, Amsterdam, Netherlands

[2]Departments of Psychological Methods, University of Amsterdam, Amsterdam, Netherlands

[3]Department of Developmental and Educational Psychology, Leiden University, Leiden, Netherlands

Correspondence: Wilma Resing, Developmental and Educational Psychology, Leiden University, LEIDEN, Wassenaarseweg 52, 2333 AK, Netherlands.

Received: April 5, 2017 Accepted: April 25, 2017 Online Published: May 7, 2017

doi:10.11114/jets.v5i6.2326 URL: https://doi.org/10.11114/jets.v5i6.2326

Abstract

In this paper we investigated the contribution of a dynamic testing procedure, including multiple graduated prompts protocols, in identifying differences in need for instruction of second grade children (N = 120) with arithmetic difficulties. The training was adaptive and prompts were provided according to one of six protocols, each focusing on a different problem solving step. Results showed that based on the number of prompts required from each protocol different patterns of problem solving could be identified, and further four profiles of instructional needs could be distinguished. The results provide starting points for individualized instruction and support the use of dynamic testing procedures in educational settings.

Keywords: dynamic testing, learning characteristics, arithmetic difficulties

1. Introduction

An important theme in educational practice is to tailor instructions to the individual needs of children. Classroom wide instructions may be effective for some children, but others profit from a more individual approach (Caffrey & Fuchs, 2007). Although individualized remediation techniques for children with arithmetic difficulties have long been developed, research findings have not often found their way to educational practice (Dowker, 2005). One reason might be that need for instruction is difficult to capture with instruments commonly used in psycho-educational practice (Glaser, 1980).

Although educational planning and consultation have always been part of the school psychologist's practice, the choice of instruments for carrying out diagnostic assessment has primarily been developed for other purposes. Intelligence tests, for example, have been developed for classification, identification of eligibility for special education services, or for clarification of learning difficulties (Elliott, 2003). These instruments focus mostly on quantitative assessment elements (scores) and on children's deficits. Information regarding learning processes, specific learning strategies and other, more qualitative learning aspects are only sparsely taken into account. Based on test results, such as outcomes of intelligence tests, strengths and weaknesses of a child may be identified; however, the results yield only minimal information that is useful for planning interventions and guiding educational practice (Elliot, Grigorenko & Resing, 2010).

Dynamic assessment procedures have been developed in response to these shortcomings. Common among these procedures is that an intervention or training phase is included, in which the child's response to instruction and feedback on cognitive tasks can be observed (Elliot et al., 2010; Lidz, 2016). In some dynamic assessment procedures the intervention has been developed from a clinical perspective, in which instructions and interventions given are guided by an experienced examiner and by the specific characteristics of responses of the child (e.g., Lidz, 2002; Tzuriel, 2013). Other procedures, such as dynamic testing, include a more structured and standardized instruction procedure in order to better compare children's test outcomes (Schlatter & Buchel, 2000; Hessels, 2000). According to several authors, using dynamic testing procedures appears to be an appropriate measure to determine a child's potential for learning and to guide classroom instructions (Fuchs, et al., 2007; Grigorenko, 2009).

A specific structured approach of dynamic testing, in which the number of prompts needed is taken as indication of a child's potential for learning, is the use of graduated prompts techniques (Campione & Brown, 1987; Resing & Elliot, 2011; Resing, Tunteler, De Jong & Bosma, 2009). In this approach interventions are provided as long as the child is not able to independently solve a problem, according to a hierarchical structured protocol. Some of these protocols include not only specific problem solving prompts but also instructions regarding metacognitive skills such as planning and checking (Campione & Brown, 1987; Hessels-Schlatter, 2002; Resing, Touw, Veerbeek, & Elliott, 2017). Most studies including graduated prompts techniques, however, have only one structured protocol through which children are led to the solution of the test problems.

The purpose of the current study was to examine the contribution of a specific graduated prompts training with not one but multiple structured protocols, to the measurement of the potential for learning of children with arithmetic difficulties and to the detection of individual differences in children's need for instruction. Protocols were developed such that each protocol consisted of a hierarchic graduated prompts sequence focusing on a specific element of the problem solving process. Campione and Brown (1987) conducted a series of studies with graduated prompts techniques during training and transfer tasks, showing differences in required prompts between groups of children with high versus average ability (Ferrara, Brown & Campione, 1986), and also between children with intellectual disabilities and typically developing children (Campione, Brown, Ferrara, Jones & Steinberg, 1985). Resing (2000) developed a learning potential test with a series of graduated prompts training sessions and found group differences between groups of typical developing, learning disabled and slow learning children in both their need for instruction and training time. In addition, a large variability in the requirement of different types of prompts, in particular metacognitive or cognitive prompts, was found between and within groups. Further, the number of prompts provided was an important predictor of school success. Other studies have shown how the use of graduated prompts techniques enabled identification of strengths and weaknesses of children between groups with different ethnic backgrounds based on the number and type of prompts children needed (Resing et al., 2009, 2017; Resing et al., 2009; Stevenson, Heiser & Resing, 2016). Within group variability was also found in a population of children with intellectual disability (Bosma & Resing, 2012).

1.1 Research Aims

To date, as far as we are aware of, studies in which graduated prompts techniques were employed, used a single structured protocol of prompts, based on task analysis (Campione & Brown, 1987; Fabio, 2005; Hessels-Schlatter, 2002; Resing, 1990; Resing & Elliott, 2011). These protocols have been effective for tasks such as solving analogical matrices or seriation problems for typically developing children and children with learning difficulties. A single graduated protocol constructed to measure the potential for learning during the completion of a seriation task (Resing, Tunteler et al., 2009), appeared suitable for typically developing children. However, children with learning difficulties, in particular children with arithmetic difficulties, are often not able to make spontaneous use of the principles and strategies taught and tend to continue with less efficient strategies (Siegler, 2003). Hence, for this population of children working on complex tasks would require an adaptation in format of the protocol. We assumed that an elaborated protocol, in which all task elements (e.g., accurate measuring, planning, organizing) were each structured with sequences of graduated prompts, would be helpful to instruct children individually. This information would presumably guide testers to help children through the solving process and to provide rich information about children's need for instruction. In particular, because children with arithmetic difficulties may, besides having difficulties with math skills, also have deficits related to specific multi-step problem-solving processes such as establishing a problem representation and developing a solution plan (Andersson, 2008; Bryant, Bryant & Hamill, 2000; Mädamürk, Kikas & Palu, 2016). We therefore developed six scenario-protocols for each step in the problem solving process of the complex version of the Seria-Think Instrument (Tzuriel, 1998; 2000). The graduated prompts included in each protocol were structured from metacognitive, to more task specific cognitive prompts to very specific modeling prompts, in which the (partial) solution (e.g. measure depth) was shown.

1.2 Research Questions

Our first hypothesis was that trained children would show greater improvement in their use of efficient strategies between pre- and posttest than non-trained children. As a measuring strategy was instructed, we expected trained children, compared to non-trained children, to show an increase in measurement behavior and a decrease in the number of insertions they performed between pre- and posttest (Resing, Tunteler et al., 2009; Tzuriel, 2000). Secondly, we expected that a) the overall amount of help children needed would decrease over the course of the series of problems and b) some scenario-protocols would be used more often during training (problem series 1-5) than others. We also focused on changes in numbers of prompts provided during the administration of the various scenario-protocols and c) expected that these would follow different patterns. Thirdly, we expected to be able to categorize the children according their need for instruction from the results of the different scenario-protocols and explored whether these categories of children would show qualitative differences such as type of prompts (metacognitive, cognitive or modeling) that were most often required or behavioral differences during training. Furthermore possible differences in demographic characteristics, cognitive functioning and achievement between the "instructional needs" categories were analyzed.

2. Method

2.1 Participants

The present sample consisted of 120 children, 39 boys and 81 girls[1], with a mean age of 8.0 years (SD= 5.9 months). While more than two-thirds of the participants were female, the distribution of boys and girls over the conditions did not differ significantly. Children, all second grade students, were recruited from 23 primary elementary schools in (inner) cities in the Netherlands. Half of the children (52%) had ethnic backgrounds other than Dutch. All children were able to understand and speak Dutch. Children were selected based upon their low achievement scores (lowest 20%) on a national norm-referenced math test (Janssen, Scheltens & Kraemer, 2005). After schools agreed to participate, written parental consent was obtained for all children. Due to missing one or more experimental sessions, the data of three children were omitted from the study.

2.2 Design

The study had a pretest-posttest control group with randomized block design, with blocking based on a short version of Raven's SPM (Raven, Raven & Court, 2000). Children in the experimental condition were administrated static pre- and posttests during which they independently solved Seria-Think problems. Between pre- and posttest children in the experimental condition received a dynamic training. Children in the control condition remained in the classroom, following regular classroom instruction. Several tasks regarding arithmetic, memory, and seriation were administrated to control for differences between groups.

2.3 Instruments

Seria-Think Instrument. The Seria-Think Instrument has been developed by Tzuriel, and was designed as a dynamic test of seriation and early math skills (i.e. estimation, measuring, counting, addition, subtraction) (Tzuriel,1998, 2000; Tzuriel & Trabelski, 2015). Tzuriel (2000) reported pre- and posttest alpha's for insertions (α =.37*;* α= .66*)* and measurements (α =.78*;* α= .85*)*. By using measurements and as few insertions as possible, children have to put cylindrical rods in a wooden cube with series of holes in various depths, to obtain increasing, equal or decreasing heights of rods. Series differ in the depth of holes and regularity in series differs as well. In the current study we used the most complex version of the Seria-Think, consisting of a wooden cube with 5 rows of 5 holes each, a set of cylindrical rods and two measuring sticks instead of the one in the original test. Instructions for the static pre - and posttest were similar to those used by Resing, Tunteler et al. (2009).

Multiprotocol- graduated prompts training. Instead of the original mediation- training with a focus on addition and subtraction skills (Tzuriel, 1998) we designed a stepwise training procedure based on graduated prompts techniques, with a focus on accurately measuring and counting as strategies to solve each series. Tzuriel & Trabelski (2015) explain the problem-solving procedure for the Seria-Think as follows; determine the depth, determine the height, compute the depth and height together, select the correct rod and complete the series. Based on our experience with the Seria-Think instrument (Resing et al., 2009) and on characteristics of children with arithmetic difficulties, we decided that, instead of by computation, the task could as well be solved by accurate measuring, planning and selecting. We therefore developed a solution procedure existing of four steps: (1) determine depth, (2) plan the preferred height, (3) select the necessary rod and (4) complete the series. Addition procedures could be avoided, because we focused on a 'reading of' the measurement sticks strategy. When the measuring stick was put in a hole, it allowed the child to read of the depth, but also the desired height. One of the instructed strategies was that the child had to grasp his or her fingers at the desired length on the measuring stick, take it out and count all the units on the stick all the way up to the desired length (position of fingers). For each of the four steps, one or more protocols with prompts were developed, based on the hierarchical graduated prompts structure developed by Resing (2000). The new protocols included not only metacognitive and cognitive prompts but also modeling prompts, in which the tester modeled the action for the child.

For step 1, scenario- protocol (A) was developed to encourage the use of the measuring stick and not to insert rods without measuring. In addition, for children who did not measure depth accurately, protocol (B) was developed to prompt counting and measuring precisely. For step 2 one protocol was designed, (C), in which children were prompted to plan the measurement of the height of the rod with the measuring stick (reading of the preferable length). For Step 3, selecting the rod, scenario protocol (D) was developed to select the rod employing the measurement stick and in addition scenario-protocol (E) was developed to organize the rods. To complete the 4th step, scenario protocol (F) was developed, in which children were prompted to backtrack when series could not be finished because of limited rods. In Table 1 examples of the graduated prompts techniques of two protocols, A and C, are presented. An overview of the

[1]The unequal number of boy and girls were the actual results of our recruitment.

four step- solution procedure and the five protocols are depicted in the flow chart in Figure 1. For one series of five holes scenario protocols A-D and F could be employed as often as necessary, whereas scenario E, organizing rods, was only allowed to use once each series. The tester provided the minimally number of prompts the child needed to independently solve the problem. The choice of employing a scenario depended on the actions of the child. Employing these scenario-protocols enabled the tester to prompt children in response to their actions, to work systematically, and to discourage children to make insertions without any measurement or planning. In addition, positive and informative feedback was given after completing a step in the solution procedure and on completion of a series.

Table 1. Example of graduated prompt sequences for Scenario A and C

SCENARIO A: Use measuring stick Use when child inserts without measuring		
Problem/action		Type*
Correct	1a. *"This rod is accidentally the right one, but remember you may measure as much as you want with this stick, but try to insert as less rods as possible without measuring*"	MC
Incorrect	1b *"Try to measure the hole/depth with the measuring stick"*	COG
Another insertion	2 . *Provide stick and say: "with this stick you first need to measure how deep the hole is"*	COG
Scenario C: Plan height of rod Use when child measured depth but inserts/selects rod arbitrarily		
All	1." *Look again carefully, what do you have to do?"*	MC
All	2. "*Wait! Remember to insert as less rods as necessary. You may measure as much as you like. Think with your measuring stick how high the rod needs to be"*	MC
Hole 2-5	3. "*Think about the depth of the hole and the height of the previous rod. Does that help you to find a rod that becomes just as long?*	COG
Hole 1	3. *The measuring stick was long enough not to sink in the hole, Find a rod that is as long as the measuring stick.*	COG
Hole 2-5	*4. "Put the measuring stick back once more in the hole). Point on the stick how much longer or shorter your new rod should be as the previous one (POINT). What length should your rod be? "*	COG
Hole 1	*4. "To make sure the rod does not disappear in the holes, you have to find a rod that is as long or longer that the measuring stick. What length should your rod be?"*	COG
All	5. "*Look: this line (Point on the measuring stick) is exactly the same height as the previous rod. What length should your rod be?"*	COG
All	6. "*Put your measuring stick back into the hole and look for the line that is at the same height as the previous rod. keep your fingers there when you take out the measuring stick. What length should your rod be?"*	COG
All	*7. Model action: , "Your rod needs to be as long as this: from here where your fingers are, to the bottom end: This long"*	MOD

* MC= metacognitive prompt; COG = cognitive prompt; MOD is modeling prompt

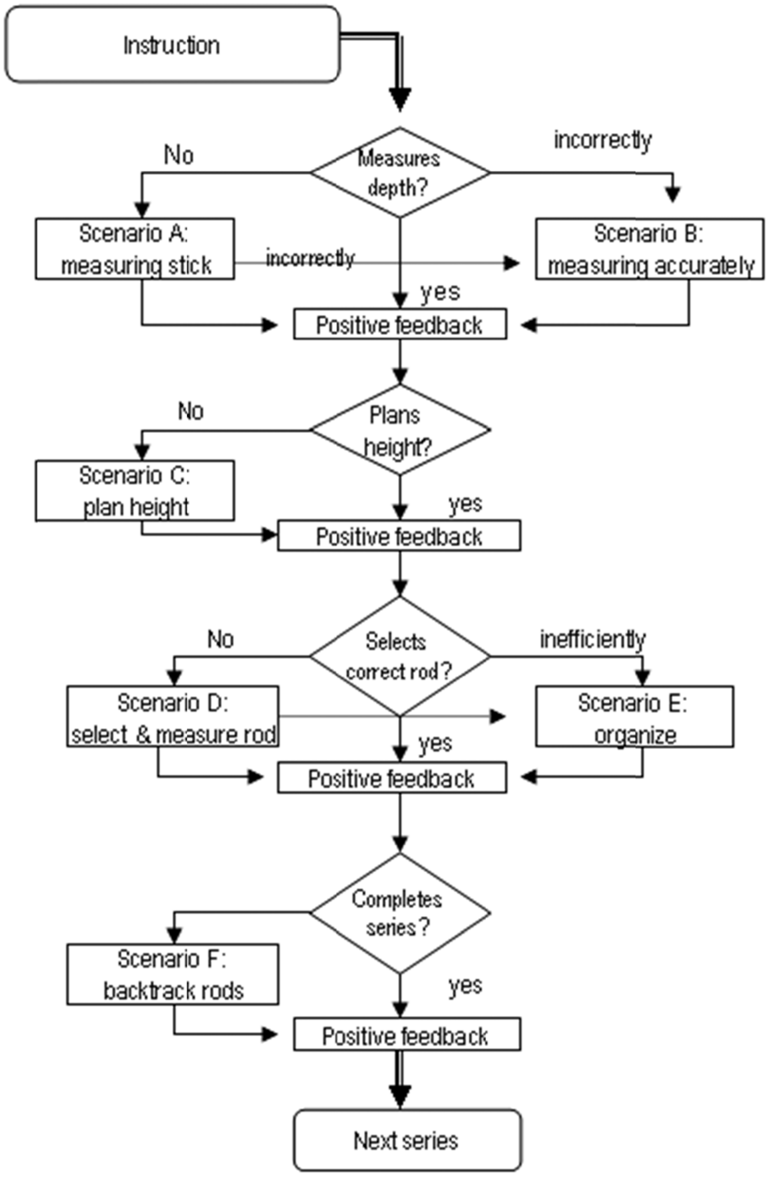

Figure 1. Flow chart of seriation-solving procedure with the scenario protocols

Behavior rating scale. The observation rating scale of 12 items (with a 6 point scale) measuring the factors openness, self-confidence, and concentration - cooperation (Resing, 1993) was completed during training.

Working memory. Correctly recalled items on the *Digit Span backwards* (WISC-III^{NL};Wechsler, 2005) and *Auditory Digit Sequencing* subtest of the Swanson-Cognitive Processing Test (S-CPT, Swanson, 1995) were scored and categorized. A recall of three numbers on Digit span and two to three items on Auditory sequencing were considered as average for this age group (Swanson & Beebe-Frankenberger, 2004).

Seriation task. The Seriation subtest Figure and Number series of the Analysis of Learning potential (Durost, Gardner & Madden, 1970), measuring logical reasoning ability, was administered.

Arithmetic test. The parts for 1st and 2nd grade level of the Dutch Didactic Age test for arithmetic/math (De Vos, 2001) was administered to capture children's arithmetic skills.

2.4 Procedure and Scoring

The Seria-think instrument (Tzuriel, 1998, 2000) was administrated individually in three weekly sessions. In all sessions five Seria-Think series had to be completed. Children were instructed to use the measuring stick as often as they liked, but to insert rods as minimally as possible; the use of the measuring stick in measuring depth was demonstrated once. During training, the child was asked to make a series of equal in height and children were prompted according to the protocols until they were able to solve series independently. The working memory, math, and seriation tasks were administrated in small groups (4-6 children). Seria-Think pre- and posttest scores were based on the numbers of insertions and measurements as well as the number of series completed correctly. Scores were directly recorded by the tester. Training sessions were recorded on video and frequencies of each type of prompt were scored for each scenario and each series by both the trainer and the first author.

3. Results

3.1 Initial Analyses

Before answering the research questions we tested whether the experimental and control group differed prior to the intervention regarding age, gender, ethnicity, and Raven, CITO math, arithmetic, memory, and seriation scores. Results of Chi-square analyses showed that the two groups did not significantly differ for gender: χ^2 (1) = .10, p = .75 or ethnic background: (χ^2 (1) = .52, p = .47. One way Anova's with condition as factor for respectively age, Raven, Cito math level, Arithmetic, Digit span backwards, Swanson memory task, and Seriation scores as dependent variables were conducted. As expected, we did not find differences, indicating that children in both conditions did not differ regarding these control variables. In Table 2 the means and standard deviations for each group are reported.

Table 2. Means and standard deviations of age, Raven, arithmetic, memory and seriation scores for the experimental and control group

	Experimental group		Control group		Total	
	Mean	SD	Mean	SD	Mean	SD
Age in months	95.58	5.870	96.75	6.041	96.17	5.961
Raven	25.07	8.497	25.92	7.866	25.50	8.161
Math CITO	1.85	.665	1.90	.730	1.87	.696
Arithmetic	32.41	9.577	31.67	10.260	32.03	9.893
Digit Span Backwards	3.64	1.200	3.65	1.313	3.65	1.253
Memory Swanson	1.71	1.232	1.85	.971	1.78	1.106
Seriation	14.08	5.187	14.80	6.014	14.45	5.607

3.2 Effectiveness of Multiple Protocols Graduated Prompts Training

First, we tested the hypothesis (1) that the trained children, compared to children in the control group, would show a decrease in the number of insertions they performed, and an increase in their measurement behavior as a consequence of the graduated prompts training. To be sure that trained children with a different ethnic background were showing comparable patterns of increases and decreases in the numbers of measures and insertions, we included ethnicity in the analysis. In Table 3 the means and standard deviations of the numbers of insertions and measurements are presented. A repeated measures (RM) analysis was performed with two dependent variables: the numbers of insertions and number of measurements. Session (pretest – posttest) was specified as within subjects factor, and condition (experimental – control group) and ethnicity as between subjects factors. There was a significant interaction effect between session and condition on the variable number of measurements, F (1,116) = 39.70, p < .001, partial η^2 =. 26; number of insertions did not show an effect (F (1,116) = 1.44, p= *ns*).

These outcomes indicate that, as expected (1), children in the experimental group improved their number of measurements significantly more from pretest to posttest than children in the control group. However, contrary to expectation and the instruction to minimize insertions this did not lead to a decrease in their number of insertions. Although the data revealed a significant main effect for ethnicity (F(2,115) = 7.80, p = .001, η^2 = .12), indicating that Dutch children on the whole showed higher mean scores at both the pretest and posttest than children with non-Dutch ethnic backgrounds, no significant interactions between ethnicity and condition (p = .79,) or ethnicity, condition and session (p = .77) were found. We therefore concluded that experimental-group but also control-group children of both ethnic groups showed comparable progressions in series completion as a consequence of dynamic testing. Therefore, in further analyses no differentiation was made regarding ethnicity.

Table 3. Means and standard deviations of number of insertions and number of measurements at pretest and posttest for the experimental and control group

	Experimental group (N=60)				*Control group (N=60)*			
	Indigenous N=31		*Ethnic Minority N= 29*		*Indigenous N=26*		*Ethnic Minority N=34*	
	pre	*post*	*pre*	*post*	*pre*	*post*	*pre*	*post*
Number insertions								
Mean	76.9	53.8	96.7	67.2	84.7	75.7	115.7	91.7
SD	25.4	23.4	34.9	27.4	35.8	34.7	50.9	41.2
Number of measurements								
Mean	7.9	21.3	7.3	19.5	7.0	10.3	6.6	7.2
SD	7.9	8.8	7.6	10.7	7.9	9.3	6.7	7.75

3.3 Patterns in Prompts Requirements

Our forthcoming research questions pertained to the patterns of change in prompt requirements during training therefore we only included children (N=60), who received the intervention. During the whole training session, children received a considerable number of prompts (M = 32.8, SD= 19.07), which varied between 1-85 prompts divided over 5 series of problems. Here we investigate three hypotheses (2a) whether the number of prompts required decreases with time and (2b) if differences could be found in the required prompts between scenario-protocols and (2c) whether different patterns in requirement of scenario prompts over the series could be detected.

A two-way single group RM analysis was conducted with the number of prompts provided per scenario-protocol as the dependent variable and Series (series 1-5) and Scenario-protocol (A-F) as within-subjects variables. A significant main effect for time was found F (4, 56) = 32.17, p < .001, partial η^2 = .69, indicating that, as expected (2a), the number of prompts decreased across series 1-5. The means and standard deviations per series are presented in Table 4.

Table 4. Mean number and standard deviation of prompts per series

		Series 1	Series 2	Series 3	Series 4	Series 5	Total
Prompts	Mean	8.30	9.70	8.23	4.75	1.90	32.88
	SD	8.87	5.64	6.88	4.71	2.89	19.07

A significant main effect for scenario-protocols was found F (5,55) = 30.86, p < .001, partial η^2 = .74, indicating that, as expected (2b) the mean number of prompts provided differed between scenario-protocols. Results of (repeated) contrasts analyses showed significant differences (p ≤ .001) between scenario-protocols A-B, B-C, C-D, and D-E. No differences were found between scenario-protocol E and F. The means and standard deviations of required prompts with each scenario are presented in Table 5. Follow-up analysis with pair-wise comparisons of scenario-protocols showed significantly different patterns between the various scenario-protocols. The most prompts were required in scenario C (p ≤ .001) compared to the other 5 scenario-protocols. This was followed by scenario A with significantly more (p ≤ .01) prompts required than B, E and F. Significantly more prompts were needed in scenario D (p ≤ .01) than E and F.

Table 5. Mean number and standard deviation of prompts per scenario-protocol

		Scenario A	Scenario B	Scenario C	Scenario D	Scenario E	Scenario F
Prompts	Mean	5.03	2.37	17.75	3.62	1.483	1.55
	SD	4.01	3.16	11.24	3.96	1.64	2.24

A significant interaction effect between scenario-protocols and series was found: F (20, 40) = 10.01, p < .001, partial η^2 = .83. This result indicates that, as expected (2c), the mean number of prompts given per scenario-protocols changed differently over the five series of problems. These change patterns are shown in Figure 2. Results of contrast analyses between series and scenario-protocol showed significant differences (p ≤ .001) between scenario-protocols B-C and C-D for series 2, 3 and 4 compared to series 5. Further, significant differences (p ≤ .05) were found between scenario-protocols A-B and D-E for series 1 and 2, compared to series 5. A difference between E and F was only found for series 1 compared to series 5 (p ≤ .05).

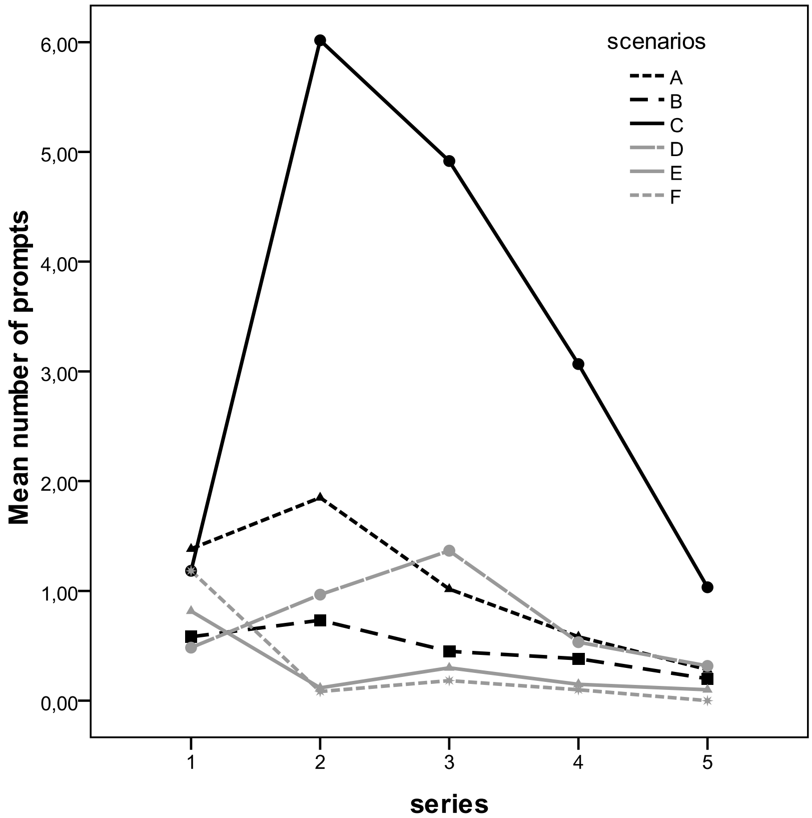

Figure 2. Mean number of prompts for each scenario over the 5 series

3.4 Profiles of Need for Instruction

A cluster analysis was conducted to investigate the hypothesis (3) that individual differences would exist between children regarding the required prompts by specific scenario-protocols. We analyzed this using the standardized scores of the total number of all prompts provided per scenario-protocols A-F, employing a two step cluster analysis procedure. The first step involved a hierarchical cluster analysis using Ward's linkage method (Bartholomew et al., 2002) and squared Euclidean distance as a measure of similarity. Inspection of the results of this cluster analysis revealed that omission of one case was necessary, as this child did not belong to any of the clusters, due too very high number of required prompts overall. Resulting agglomeration coefficients displayed a notable increase, and we therefore used a four cluster solution. In the second step we employed a k-means cluster analysis, specifying the four-cluster solution using simple Euclidian distance as the similarity measure. The resulting clusters included 27 children in Cluster 1, 13 in Cluster 2, 10 in Cluster 3 and 9 in Cluster 4.

We conducted a MANOVA with total number of provided prompts per scenario (A-F) as dependent variables and cluster (1-4) as between subjects factor to confirm that the four profiles significantly differed on children's need for instruction of the scenario protocols. The results revealed, as expected (3), that children in the four clusters significantly differed in prompts provided per scenario-protocol: *Wilks'Λ* = .057, $F(18, 141) = 13.76$, $p < .001$, partial η^2 =. 061. Additionally, results of paired comparisons (Bonferroni correction used) revealed significant differences between clusters for each scenario ($p < .001$) (see Table 6). Children differed by cluster in the provided numbers of prompts to use the measurement stick (scenario A). Children in clusters 1 and 3 required fewer measurement prompts (Scenario A) than children in clusters 2 and cluster 4. Children in cluster 3 required more prompts to increase accuracy (scenario B) compared to children in clusters 1 and 2. Children in clusters 1 and 2 were provided fewer prompts to measure height than children in clusters 3 and 4. Further, children in clusters 1, 2 and 4 were given significantly fewer prompts in selecting the necessary rod (scenario D) than children in cluster 3. Further, children in cluster 4 were provided significantly more prompts to organize the rods (scenario E) than children in clusters 1, 2 and 3 and were given more prompts to backtrack (scenario G) than children in clusters 1 and 2. Hence, we categorized the clusters as (1) General prompts, (2) Measuring prompts, (3) Accuracy prompts and (4) Organizing prompts. The resultant profiles are shown in Figure 3.

Table 6. Significant Anova results and significant difference tests for scenario by clusters

Scenario-protocol	F (3,55)	η^2	Significance	Compare clusters	Mean differences*	Confidence interval Lower	Upper
A	23.69	.56	< .001				
				1-2	-6,44*	-8,98	-3,91
				1-4	-6,22*	-9,11	-3,33
				3-2	-5,90*	-9,06	-2,74
				3-4	-5,68*	-9,12	-2,23
B	7.05	.28	< .001				
				1-3	-2,96*	-5,10	-,82
				2-3	-3,93*	-6,36	-1,50
C	16.11	.47	< .001				
				1-3	-16.51*	-24.66	-8.35
				1-4	-16.96*	-25.44	-8.49
				2-3	-10.56*	-19.82	-1.30
				2-4	-11.02*	-20.57	-1.47
D	15.04	.45	< .001				
				1-3	-7.44*	-10.52	-4.37
				2-3	-5.77*	-9.26	-2.28
				3-4	4.44*	.63	8.26
E	12.08	.40	< .001				
				1-4	-2.5185*	-3.8415	-1.1955
				2-4	-3.1282*	-4.6187	-1.6377
				3-4	-2.3667*	-3.9460	-.7874
F	15.35	.46	< .001				
				1-3	-2.98*	-4.71	-1.25
				1-4	-3.89*	-5.69	-2.09
				2-4	-2.73*	-4.75	-.70

* $p < .05$

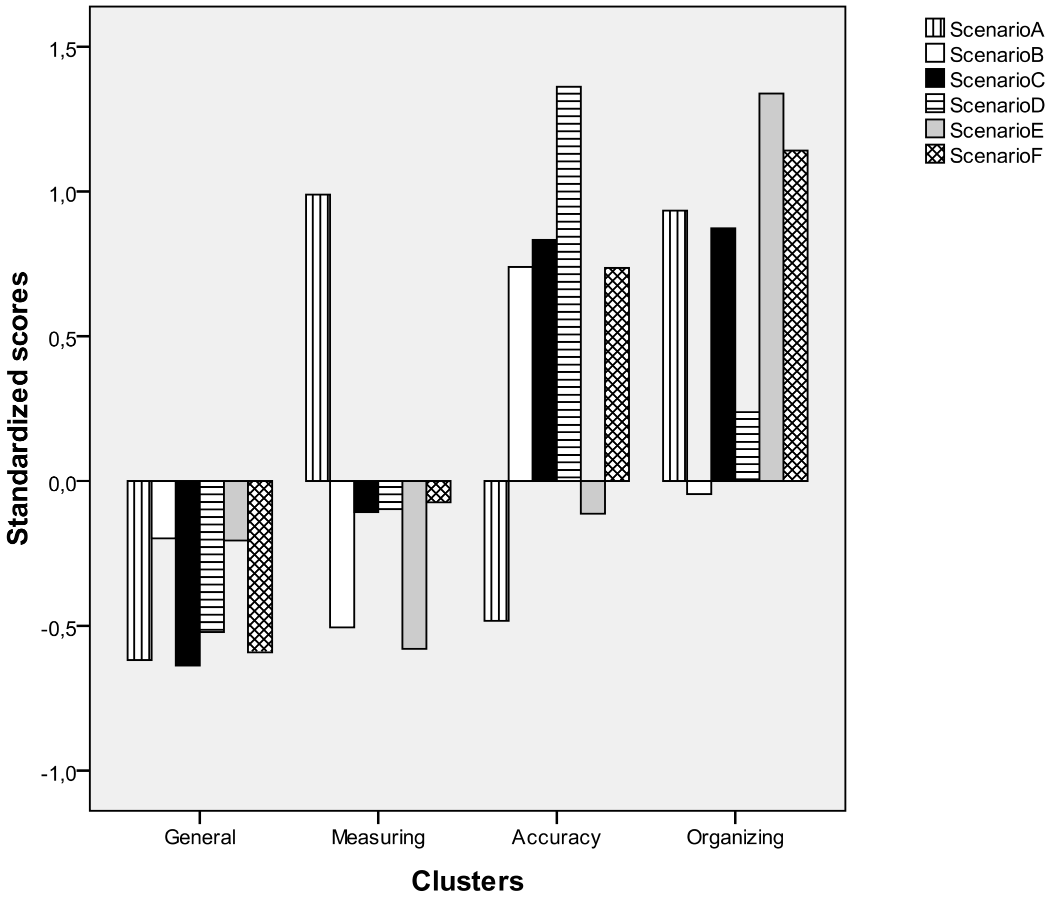

Figure 3. General, Measuring, Accuracy and Organizing prompts groups resulting from cluster analysis

3.5 Differences between Profiles of Instructional Needs

We looked at the differences in types of prompts required per profile using a MANOVA with total metacognitive, cognitive and modeling prompts as dependent variables and profiles as between-subjects factor. Here we found a significant effect for profile: *Wilks'Λ* = .35, $F(9, 129.14) = 7.74$, $p < .001$, partial $\eta^2 = .29$. Results of ANOVA's revealed significant results for metacognitive prompts ($F(3, 55) = 8.52$, $p < .001$, partial $\eta^2 =. 32$), and cognitive prompts ($F(3, 55) = 24.37$, $p < .001$, partial $\eta^2 = .57$) and modeling prompts ($F(3, 55) = 3.11$, $p \leq .05$, partial $\eta^2 =.14$). Results of contrast analyses showed how children in the General profile required significant fewer metacognitive prompts ($p \leq .01$) and cognitive prompts ($p \leq .01$) compared to children in the other profiles. Children in the Measuring profile needed significant fewer cognitive prompts ($p \leq .001$) compared to children with an Accurate or Organizing instruction profile, but no differences in required type of prompts between children in the Accurate or Organizing profiles were found.

In order to answer the question of how individual differences in required prompts would relate to the posttest results of the dynamic test and to measures of ability and achievement, we examined possible differences in cognitive functioning and achievement between the children in the four instruction profiles. Results of a MANOVA indicated no differences regarding age, Raven, Cito math level, Arithmetic, Digit span backwards, Swanson memory task, and Seriation scores during pre and posttest. Additional Kruskal-Wallis analyses indicated that the children in the profiles did not differ by gender and ethnicity as well. A significant difference between the children allocated to the four profiles was found for the initial CITO- math achievement test: X^2 (3) = 9. 12, $p = .028$. Follow-up analyses with the Mann Whitney U test showed that the children in the Accuracy profile had significantly lower levels on the CITO-math achievement task than those with a general profile ($z = -2.95$, $p = .003$).

Further differences among the instruction profiles were analyzed regarding children's behaviors observed during the training session. Results of a MANOVA with the scores of the three behavior scales (Confidence, Openness, and Concentration & Cooperation) as dependent variables and instruction profiles as between-subjects factor demonstrated significant differences in behavior among the children in the four profiles: *Wilks' Λ* = .535, F (3, 9) = 4.20, $p <.001$, partial $\eta^2 = .19$). Results of ANOVAs revealed significant differences for all three observed behaviors between the four profiles of children: observed Self Confidence ($p = .004$), observed Openness, $p =.017$ and observed Concentration & Cooperation ($p = .004$). Children with a Measuring profile were significantly more self-confident and had higher spontaneous and candid behaviors, compared to children with an Organizing profile, and greater self-confidence compared to children in the Accuracy cluster. Further, children with a General profile had significantly higher scores on observed behaviors of concentration and cooperativeness, than children in the Measuring and Organizing profile.

4. Discussion

This study sought to investigate the contribution of a dynamic testing procedure, including multiple protocols, in identifying differences in the need for instruction of second grade children with arithmetic difficulties. A first conclusion of the present study is that, as a consequence of dynamic testing, including the training we developed, children showed significantly more sophisticated problem solving strategies on the seriation task than untrained children. Trained children increased their number of measurement activities with the measuring stick, which is in line with earlier findings with another version of the Seria-Think instrument (Resing, Tunteler et al., 2009; Tzuriel, 2000). Although the number of insertions did not significantly decrease this could be due on the one hand to the relatively complex problems, as we used the five by five version of the instrument. On the other hand comparable results were found in previous studies (Resing, Tunteler et al., 2009). Also in line with these studies scorings-patterns during testing between Dutch and children with an ethnic background did not differ. Therefore, we have treated the group as a whole.

Regarding the use of the newly developed multiple protocols that had to be followed during the graduated prompts training, we found that these protocols, as expected, provided information about both the types and numbers of prompts in the feedback children were provided, which is in agreement with previous studies with a graduated prompts training (Bosma & Resing 2012; Jeltova et al., 2007; Fabio 2005). Moreover, children's needs of different prompts at different times in the solution process could be detected clearly. The use of these multiple protocols enabled us to identify different patterns of problem solving, by the number of prompts needed at each step of the problem solving procedure of each individual child. Prompts were often required during the first two steps of the problem solving process, in which children were prompted to use the measuring stick in measuring depth and planning height. Without measurement tool use it is not possible to solve the series. Furthermore, tool use has often been considered a demonstration of planning and goal directed behavior (Keen, 2011; Miller, 1989). It appears that understanding the necessity of using the measuring stick, enabled at least part of the trained children to solve further series without many additional prompts. During the third and fourth step of the problem solving process, a substantial number of prompts were also needed. Although knowledge regarding measuring was necessary to solve a series, systematic, organized and accurate working appeared to be required as well.

Despite the complexity of the seriation task we administered in our experiment, children with arithmetic difficulties showed more advanced and independent problem solving behavior during the last two seriation problems. As expected, the number of prompts provided decreased considerably over the five series and these findings correspond to findings regarding typically developing children solving comparable tasks (e.g., Resing, Tunteler et al., 2009).

Besides detailed information regarding instruction provided at each step of the problem solving procedure, the multiple protocols enabled us to identify groups of children with different profiles of "instructional needs". The largest group we identified, the General instruction profile, needed relatively few prompts per scenario protocol. A second group of children, the Measuring instruction profile, was characterized by their need for prompts regarding measuring stick use at the start of the problem solving procedure, which was, as stated above, an essential element in solving the problem. The third and fourth group of children both needed relatively more prompts. The third profile, Accuracy instruction, needed besides planning height, specific prompts in accurate measuring and selecting, with relatively fewer measuring prompts. Children in the fourth profile, Organizing instruction, primarily required prompts regarding organizing, planning and measuring.

Children categorized in the General instruction profile, needed fewer prompts, both metacognitive and cognitive prompts, compared to children in the other three "instructional needs" profiles. In addition, children with a Measuring instruction profile needed fewer cognitive prompts compared to children with an Accuracy or Organizing instructional profile. Children in the various instruction-profiles did not differ with respect to gender, ethnicity, their levels of general cognitive functioning, memory and seriation. They did however differ in arithmetic scores where children with an Accuracy instruction -profile had lower arithmetic performance than children categorized by the General instruction -profile. Inaccuracy is a repeatedly reported behavior of children with arithmetic difficulties (e.g. Bryant et al., 2000). Other differences were found regarding observed behaviors during training. Children in the Measuring profile showed more spontaneous activities during training compared to children categorized in the Accuracy and Organizing profiles; and showed greater self-confidence compared to peers in the Organizing profile. Children in the General instruction profile showed more concentration and cooperation behaviors compared to children in the Measuring -and Organizing profiles. These differences we found in instructional needs and work-attitudes could be helpful information regarding classroom recommendations. Further studies, however, have to be conducted to generalize our findings to larger populations and to understand these specific relations into more detail.

Our findings do have several implications concerning interventions and instructions for children with arithmetic difficulties. At first sight these children all achieve low scores and may demonstrate similar difficulties in understanding and solving (math) problems. In solving novel problems however, they appear to profit from different amounts and types of instruction. We therefore suggest that the use of a dynamic test procedure including graduated prompts scenarios could be a useful part of a school psychologist's repertoire. The dynamic test format including graduated prompts techniques clearly provides additional information regarding children's learning processes during complex cognitive task solving, which goes beyond defining the child's level of math ability. The test outcomes also provide detailed insights in instructions a child does or does not profit from. Both aspects of dynamic test information could be helpful for school psychologists in providing recommendations for the classroom teacher.

Furthermore, knowing that these different needs for instruction exist, may help classroom teachers recognize these needs and take a proactive approach in planning and adapting their instructions. The group of children with Accuracy or Organizing profiles appeared particularly in need of additional attention and instruction regarding general task and self-monitoring behaviors, which are essential elements in achievement (Bryant et. al., 2000; Swanson & Beebe-Frankenberger, 2004).

Although the multiple protocols appeared very informative regarding children's instructional behavior during testing, they were rather difficult for testers to learn, apply and score. In response to each action of a child the tester had to choose the right protocol, and prompts were sometimes helpful for some but not for other children. A computerized training, such as a tangible table (Henning, Verhaegh & Resing 2011 ; Resing & Elliott, 2011: Resing et al., 2017), would certainly be helpful in supporting the administration and scoring of the necessary prompts. Henning et al., 2011 showed, for example, the effectiveness of a short adaptive scaffolding procedure on an electronic console. For the moment, however, responding to all actions and behaviors of the child would remain a challenge.

To conclude, with a one session graduated prompts training, including multiple protocols, we were able to identify specific instructional needs within a group of young children with severe arithmetic difficulties, and specify at which step in the problem solving process different children needed help. Applying dynamic testing procedures as described in this study would certainly help to tailor instructions to the needs of students. Since this procedure fits very well with the current educational focus of needs based assessment, planning interventions and response to interventions (Brown-Chidsey & Andren, 2012; Fuchs et al., 2011; Pameijer, 2017), we strongly recommend adding dynamic test procedures more often to the school psychologist's assessment repertoire. A screening including a dynamic test also highly improves the prediction which children need (early) interventions regarding math achievement (Fuchs et al., 2011). Dynamic assessment may even become a first choice instrument in psycho-educational testing practices where the main requests are to provide the most suitable interventions and instructions for each individual child.

Acknowledgements

We thank J. van Wingerden for helping us code the data and E. Tunteler for her feedback on earlier drafts of this article.

References

Andersson, U. (2008). Mathematical competencies in children with different types of learning difficulties. *Journal of Educational Psychology*, *100*, 48-66. https://doi.org/10.1037/0022-0663.100.1.48

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2002). *The analysis and interpretation of multivariate data for social scientists Text in statistical science.* US: Chapman & Hall/CRC.

Bosma T., & Resing, W. C. M. (2012). Need for instruction: Dynamic testing in special education. *European journal of Special Needs Education, 27,* 1-19. https://doi.org/10.1080/08856257.2011.613599

Brown-Chidsey, R., & Andren, K. J. (Ed.). (2012). *Assessment for intervention: A problem solving approach*. New York: Guilford Press.

Bryant, D. P., Bryant, B. R., & Hamill, D. D. (2000). Characteristic behaviors of students with LD who have teacher-identified math weaknesses. *Journal of Learning Disabilities, 33*, 168-177. https://doi.org/10.1177/002221940003300205

Caffrey, E., & Fuchs, D. (2007). Differences in performance between students with learning disabilities and mild mental retardation: Implications for categorical instruction. *Learning Disabilities Research & Practice, 22*, 119-128. https://doi.org/10.1111/j.1540-5826.2007.00236.x

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82-115). New York: The Guilford Press.

Campione, J. C., Brown, A. L., Ferrara, R. A., Jones, R. S., & Steinberg, E. (1985). Breakdowns in flexible use of information: Intelligence–related differences in transfer following equivalent learning performance. *Intelligence, 9,* 297-315. https://doi.org/10.1016/0160-2896(85)90017-0

De Vos, T. (2001) *Didactic learning Equivalent test Atithmatic/ Math: manual and copying-sheets*–Leeuwarden, NL: Eduforce.

Dowker, A. (2005). Early identification and intervention for students with mathematics difficulties. *Journal of Learning Disabilities, 38*, 324-332. https://doi.org/10.1177/00222194050380040801

Durost, W. N., Gardner, E. F., & Madden, R. (1970). *Analysis of learning potential.* New York: Harcourt, Brace & World.

Elliott J. G., Grigorenko E. L., & Resing W. C. M. (2010). Dynamic assessment. In: P. Peterson, E. Baker, B. McGaw, (Ed.), International encyclopedia of education (vol. 3), pp. 220-225. Oxford: Elsevier. https://doi.org/10.1016/b978-0-08-044894-7.00311-0

Elliott, J. G. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review, 55*, 2003. https://doi.org/10.1080/00131910303253

Fabio, R. A. (2005). Dynamic assessment of intelligence is a better reply to adaptive behavior and cognitive plasticity. *Journal of General Psychology, 132*, 41. https://doi.org/10.3200/genp.132.1.41-66

Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning rules: studies of proximal development. *Child Development, 57*, 1087-1099. https://doi.org/10.2307/1130433

Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities*. https://doi.org/10.1177/004005990703900508

Fuchs, D., Fuchs, L. S., Compton, D. L., Bouton, B., Caffrey, E., & Hill, L. (2007). Dynamic assessment as Response to Intervention. A scripted protocol to identify young at-risk readers. *Teaching Exceptional Children, 39*, 58-63. https://doi.org/10.1177/004005990703900508

Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Hamlett, C. L., & Seethaler, P. M. (2011). Two-stage screening for math problem-solving difficulty using dynamic assessment of algebraic learning. *Journal of learning disabilities*, *44*(4), 372-380. https://doi.org/10.1177/0022219411407867

Glaser, R. (1981). The future of testing. A research agenda for cognitive psychology and psychometrics. *American Psychologist, 36,* 923-936. https://doi.org/10.1037/0003-066x.36.9.923

Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities, 42*, 111-132. https://doi.org/10.1177/0022219408326207

Henning, J. R. Verhaegh, J., & Resing, W. C. M. (2011). Creating an individualized learning situation using scaffolding in a tangible electronic series completion task. *Educational and Child Psychology, 28*(2), 85-100.

Hessels, M. G. P. (2000). The Learning potential test for Ethnic Minorities: A tool for standardized assessment of children in kindergarten and the first years of primary school. In C.S. Lidz & J. Elliott, *Dynamic Assessment: Prevailing models and applications* (pp 109-131). New York: Elsevier.

Janssen, J., Scheltens, F., & Kraemer, J. (2005). *Student and education Monitoring System. Mathematics Grade 2. Teachers manual*. Arnhem, NL: CITO.

Keen, R. (2011). The development of problem solving in young children: a critical cognitive skill. *Annual Review of Psychology.* DOI: 10.1146/annurev.psych.031809.130730

Lidz, C. S. (2002) Mediated learning experience (MLE) as a basis for an alternative approach to assessment. *School Psychology International, 23,* 68-84. https://doi.org/10.1177/0143034302023001731

Lidz, C. S., & Elliott, J. G. (Eds.) (2000). Advances in cognition and educational practice (series editor: J. S. Carlson) Vol.6. *Dynamic assessment: Prevailing models and applications* (pp 3-13). Oxford: Elsevier Inc.

Lidz, C.S. (2016) Around the World with the Application of Cognitive Functions Scale. Unpublished manuscript, uploaded to researchgate.net, March 28, 2016

Mädamürk, K., Kikas, E., & Palu, A. (2016). Developmental trajectories of calculation and word problem solving from third to fifth grade. *Learning and Individual Differences*, *49*, 151-161. https://doi.org/10.1016/j.lindif.2016.06.007

Miller, K. F. (1989). Measurement as a tool for thought: the role of measuring procedures in children's understanding of quantitative invariance. *Developmental Psychology, 25,* 589-600. https://doi.org/10.1037//0012-1649.25.4.589

Pameijer, N. K. (2017). Assessment and intervention. Bridging the gap with a practice-based model*. Educational and Child Psychology, 34,* 66-82.

Raven, J. C., Court, J. H., & Raven, J. (1979). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* London: H. K. Lewis & Co.

Resing, W. C. M. (2000). Assessing the Learning potential for Inductive Reasoning (LIR) in young children. In C. S Lidz & J. G. Elliott (Eds*.), Advances in cognition and educational practice:* Vol. 6: *Dynamic assessment: Prevailing models and applications* (pp. 229-262). New York: Elsevier.

Resing, W. C. M., & Elliott, J. G. (2011). Dynamic testing with tangible electronics: Measuring children's change in strategy use with a series completion task. *British Journal of Educational Psychology*. https://doi.org/10.1348/2044-8279.002006

Resing, W. C. M., Touw, K. W. J., Veerbeek, J., & Elliott, J. G. (2016). Progress in the inductive strategy-use of children from different ethnic backgrounds: a study employing dynamic testing. Educational Psychology, 37(2), 173–191. https://doi:10.1080/01443410.2016.1164300

Resing, W. C. M., Tunteler, E., De Jong, F. J., & Bosma, T. (2009). Dynamic testing: Individual differences and learning patterns in indigenous and ethnic minority children. *Learning and Individual Differences, 19,* 445-450. https://doi.org/10.1016/j.lindif.2009.03.006

Schlatter, C., & Büchel, F. P. (2000). Detecting reasoning abilities in persons with moderate mental retardation: The Analogical Reasoning Learning Test (ARLT). In C.S. Lidz & J. Elliott (Eds.) *Dynamic Assessment: Prevailing Models and Applications* (pp.155–186). New York: Elsevier Sciences Inc.

Siegler, R. S. (2003). Implications of cognitive science research for mathematics education. In: Kilpatrick, J., Martin, W. B., & Schifter, D. E. (Eds.), *A research companion to principles and standards for school mathematics (pp. 219-233).* Reston, VA: National Council of Teachers of Mathematics.

Speece, D. L., Cooper, D. H., & Kibler, J. M. (1990). Dynamic assessment, individual differences, and academic achievement. *Learning and Individual Differences, 2*, 113-127. https://doi.org/10.1016/1041-6080(90)90019-d

Stevenson, C. E., Heiser, W. J., & Resing, W. C. M. (2016). Dynamic testing: Assessing cognitive potential of children with culturally diverse backgrounds. *Learning and Individual Differences, 47*(2016), 27-36. https://doi.org/10.1016/j.lindif.2015.12.025

Swanson, H. L. (1995). *S-Cognitive Processing Test*. Austin, TX: Pro-Ed.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 96,* 471-491. https://doi.org/10.1037/0022-0663.96.3.471

Tzuriel, D. (1998). *Seria-Think Instrument. Instruction manual*. Bar-Ilan University, ISR: School of Education.

Tzuriel, D. (2000). The Seria-Think instrument: A novel measure for assessment and intervention in seriational-computational domain. *School Psychology International, 20*, 173-190. https://doi.org/10.1177/0143034300212005

Tzuriel, D. (2013). Dynamic Assessment of Learning Potential. In: M. Mo Ching Mok (ed.), *Self-directed learning oriented assessments in the Asia-Pacific, education in the Asia-Pacific region: Issues, concerns and prospects 18*, https//doi:10.1007/978-94-007-4507-0_13

Tzuriel, D., & Trabelsi, G. (2015). The effects of the Seria-Think Program (STP) on planning, self-regulation and math achievements among Grade 3 Children with attention deficit hyperactivity disorder (ADHD), 345-367. In T. Papadopoulos, R.K. Parrila, J R. Kirby (eds). *Cognition, intelligence, and achievement: A tribute to JP Das*,. Amsterdam: Elsevier . http://dx.doi.org/10.1016/B978-0-12-410388-7.00017-8

Wechsler, D. (2005). Wechsler Intelligence Scale for children III. Dutch version. Lisse: Harcourt.