



Universiteit
Leiden
The Netherlands

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

Touw, K.W.J.

Citation

Touw, K. W. J. (2020, September 17). *Computerised dynamic testing: An assessment approach that tailors to children's instructional needs*. Retrieved from <https://hdl.handle.net/1887/136755>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/136755>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/136755> holds various files of this Leiden University dissertation.

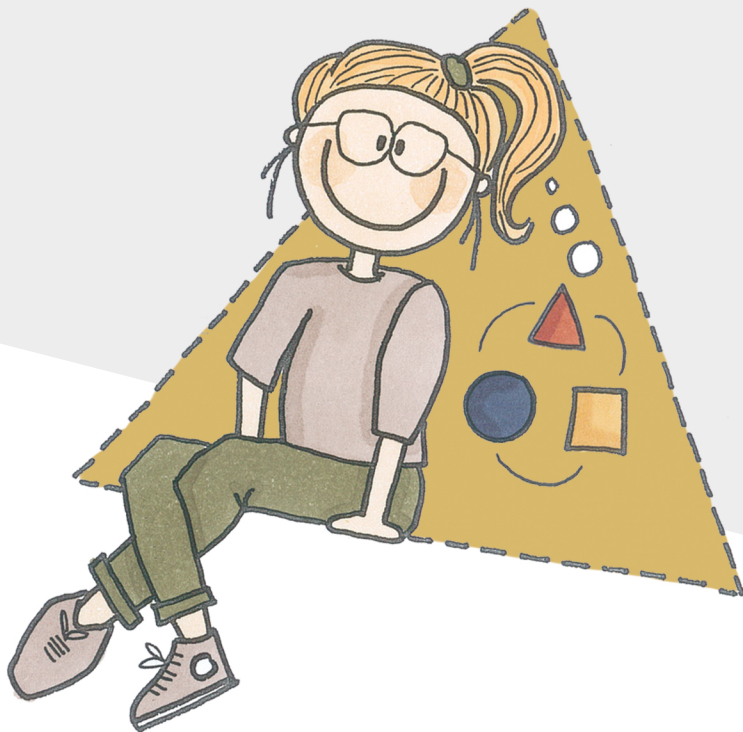
Author: Touw, K.W.J.

Title: Computerised Dynamic Testing: An assessment approach that tailors to children's instructional needs

Issue Date: 2020-09-17

Chapter 5

CHILDREN'S PROGRESS IN SOLVING FIGURAL
ANALOGIES: ARE OUTCOMES OF DYNAMIC
TESTING HELPFUL FOR TEACHERS?



KIRSTEN W. J. TOUW
BART VOGELAAR
ROEL VERDEL
MEREL BAKKER
WILMA C. M. RESING

Touw, K. W. J., Vogelaar, B., Verdel, R., Bakker, M., & Resing W. C. M. (2017). Children's progress in solving figural analogies: Are outcomes of dynamic testing helpful for teachers? *Educational and Child Psychology*, 34(1), 21-38.

Abstract

Background: Static test results reflect what children have learned up to the moment of testing. Dynamic tests, on the other hand, aim to examine children's potential for learning. The information obtained during dynamic testing has the potential to be helpful for teachers. However, studies examining the opinion of teachers regarding the usefulness of dynamic testing outcomes are scarce.

Aim: The main aim of the present study was to investigate the advantages of dynamic testing for education. We aimed to provide insight into the potential for learning of first grade children in weak and strong functioning groups, as indicated by the teacher. The present study also sought to provide teachers with dynamic test results to enable individualised educational practices for the participating children. Rationale/approach: Thirty children, either weakly or strongly performing in the classroom were tested dynamically by means of a pre-test, two training sessions, and a post-test. This procedure enabled a detailed examination of children's learning processes while solving figural analogies. The practical value of dynamic testing was examined by asking teachers and educational consultants to what extent the dynamic test outcomes could contribute to instruction plans.

Findings/conclusion: The children in the weak functioning group scored, on average, lower on the figural analogies test than children in the strong functioning group. Both groups, however, benefited similarly from the training. Individual differences in reasoning progress and the type and number of prompts needed during training were also found in both groups. Children's scholastic achievement scores and teacher ratings of school performance were both related to an equal extent to the dynamic and static test measures of the dynamic test. Lastly, the evaluations of the teachers supported the view that dynamic test results can have practical value for education.

5.1 Introduction

Assessment procedures are used increasingly frequently in education in order to take stock of children's development. On the basis of reports and consultations with educational psychologists and/or educational advisory services, teachers can monitor which children have difficulty with a specific course or skills within a certain domain. They can also take into account children's progress as to their scholastic performance in comparison with the education they have received. Individual test scores can be compared with the mean score of children of a certain age, and converted to a mark. Sometimes, ipsative assessment is sufficient, for instance as regards student tracking systems that enables comparison of results of individual learners against their previous test results (e.g., Bourke & Mentis, 2014). Sometimes additional assessment is necessary, for example as regards questions within the framework of inclusive education, when individual students' level of general cognitive abilities, intelligence or potential for learning is also taken into account.

Static test results, which are single-session tests that do not include any feedback or help, provide an indication of what a child has learned up to the moment of testing. These test scores, however, do not optimally reflect the *ability to learn* (Haywood & Lidz, 2007; Resing, 2000; Robinson-Zanartu & Carlson, 2013; Sternberg & Grigorenko, 2002; Tzuriel, 2013), and provide a very limited insight into the process involved in *how children learn* (Grigorenko, 2009). *Dynamic tests*, however, aim to provide an overview of a child's potential for learning. During a dynamic test, help is offered that, depending on the specific test, can be adaptive. This means that the help can be tailored to the individual child's level. The primary aims of dynamic testing research have been to examine or reveal improvements in cognitive abilities following training between test session(s), to observe and measure behaviour related to the individual's potential for learning, and to gain insight into learning processes at the moment assessment takes place (Elliott, Grigorenko, & Resing, 2010). Research has shown that static test results can underestimate children's cognitive abilities, for example in the case of children with learning or language difficulties, and children with different cultural backgrounds (e.g., Hessels, 2000). In addition, static tests do not provide sufficient information as to *how* a child learns, and from what *type of instruction* this child benefits most. Some proponents of dynamic testing argue that such information would be necessary in order to potentially guide teachers or inform recommendations about additional forms of intervention in the classroom (Grigorenko, 2009; Jeltova et al., 2011). Attempts to measure learning processes while children are being trained and tested, in combination with informing teachers about these assessment outcomes, however, are scarce.

Despite their – potential – advantages for education, dynamic tests are not yet widely used (Elliott, 2003; Elliott & Resing, 2015), primarily due to the fact that conducting these tests often takes extra time. The aim of the current study was therefore to provide insight into the advantages of dynamic testing for education, so that investing extra time in assessment can be justified.

A few years ago, a dynamic test study was conducted in cooperation with a large educational advisory service (Stevenson et al., 2011). This dynamic test had a test-training-retest structure, and was developed with a graduated prompts model, by which the experimenter gradually guides the learning processes and strategies by giving the child training based on step-by-step solution procedures (e.g., Campione & Brown, 1987; Fabio, 2005; Resing, 2000; Resing & Elliott, 2011). During training, help, by means of a scaffolding procedure, is only provided when children need it. The amount of help, in addition to the progression from pre-test to post-test, is taken as a measure of the potential for learning. The test is assumed to measure the potential for solving reasoning tasks, by means of looking into children's analogical reasoning skills. Analogical reasoning, a form of inductive reasoning, is considered critical for both cognitive development, and learning in a school context (Goswami, 2002). Inductive reasoning tasks, such as analogical reasoning, categorisation, and seriation, all require a limited set of rule finding processes that can be induced by searching for similarities and differences between the objects, and/or in the relations between the task objects (Klauer, 2014; Klauer & Phye, 2008; Perret, 2015). Test items such as analogies, seriations, and matrices are often central in measuring intelligence, and they are, therefore, assumed to be highly suitable for providing a general estimation of children's potential for learning (Sternberg & Grigorenko, 2002). In previous studies, analyses of the processes involved in solving analogy tasks were, therefore, used to develop the training intervention of the dynamic test further (Resing, 1993; Sternberg & Gardner, 1983; Stevenson, Resing, & Froma, 2009).

This former study by Stevenson and colleagues (2011) was primarily aimed at answering the question as to whether dynamic test scores derived from solving analogy tasks could be used in predicting young children's scholastic achievement. The dynamic test scores were found to significantly contribute to the prediction of young children's scholastic achievement in grade 1 of regular education. Furthermore, the dynamic test scores provided more qualitative information as to the number and the type of prompts that children needed, and as to the progression in reasoning ability as a consequence of an intervention. This qualitative information was valued highly by teachers and educational consultants, and would not have been available if static tests only had been used.

Aims of the present study

The aim of the present study was to provide more insight into the potential for learning of children from weak and strong functioning groups, as indicated by the teacher. In addition, the present study aimed at providing suitable feedback to teachers to enable individualised educational practices for the children participating in the study. Using the instrument used in the former study of Stevenson and colleagues (2011) enabled us to compare the data of the children in the present study with the norms obtained in the former study, so that the children's potential for learning could be estimated.

The present study was aimed at comparing the dynamic test results of two groups of children from grade 1. We expected that the group of children categorised as weak functioning by their teacher would show more progression in solving analogy tasks after training than the group of children categorised as strong functioning by their teacher, which has been found in previous research, for example in a study by Stevenson, Heiser, and Resing (2013). In addition, we expected significant individual variation. We also expected that the children categorised as strong functioning would need less explicit help during the training sessions than the children categorised as weak functioning, although, here, we also expected to find substantial individual variation. Moreover, we explored the relationship between, on the one hand, the teacher rating of children's scholastic achievement and the actual scholastic achievement as measured by the student tracking system (Cito, 2012), and, on the other hand, the static and dynamic measures obtained through the test. We expected that the teacher rating as to the potential for learning and the children's scholastic achievement scores, would show a stronger relationship with the dynamic test measures than the static pre-test scores of the dynamic test (AnimaLogica). We also expected that the children's actual scholastic achievement results would show a stronger relationship with the dynamic test measures than the static pre-test scores (Resing, 2000).

In order to further test the practical value of the information obtained by means of the dynamic testing procedure, both teachers and educational consultants were asked, after the children had been tested dynamically, to what extent static and/or dynamic test information could contribute to instruction and action plans. It was expected that the dynamic test results could enable insight into learning ability, and contribute to education, both as regards prediction of future achievements, and instruction-based assessment (Bosma & Resing, 2008; Brown-Chidsey & Steege, 2005; Pameijer, 2006; 2016).

5.2 Method

Participants

Eighty children participated in the present study, who attended one of three classes of grade 1 of one school offering regular primary education. The children had a mean age of 6 years and 10 months. All parents gave their written consent prior to participation in the study.

Design and procedure

Prior to the study, for each child, the teachers in question filled out a questionnaire regarding scholastic achievement and potential for learning. On the basis of these questionnaires, in each class, the 25% lowest and the 25% highest performing children were selected for participation. The overall estimation of the teacher rating was taken into account, and the teacher rating as to the potential for learning of the children in question was the decisive factor. Additional data collected as part of the

present study included the Cito scholastic achievement scores for maths, technical reading, and spelling. To answer our question concerning the potential for learning of children from the weak and strong functioning groups, as indicated by the teacher, 30 students (16 weak functioning and 14 strong functioning) participated in the present study.

All children, both the weak and strong functioning children, were tested individually and dynamically by means of a sandwich design (Sternberg & Grigorenko, 2002), consisting of a pre-test (computer), training 1, training 2, and a post-test (computer). Prior to the dynamic test, a mouse task was conducted to check whether the children had sufficient motor skills to carry out the computer task correctly, and understood how the computer mouse functioned (dragging and clicking). A week after the pre-test the first training session was conducted, in which concrete material was used. Another week later, the second training session was conducted, and yet another week later the post-test. The pre-test and post-test took approximately 20 minutes, and each training session approximately 30 minutes. After conducting the dynamic test, reports were drawn up, and evaluated by the educational consultant with the teachers.

Instruments

School achievement measures

Teacher rating questionnaire. The questionnaire consisted of eight questions (five-point scale). The teacher was asked to rate each child's scholastic and cognitive achievement, relative to children of the same age, including children outside the class they were teaching. The following eight domains were rated: Maths, Language, Technical reading, Spelling, Writing, Reasoning, Potential for learning, and Overall rating of scholastic achievement. The teacher was asked to divide the children's achievement into five categories, of which '1' indicated the 20 per cent 'weakest performing' children, and '5' the 20 per cent 'strongest performing' children.

LOVS student tracking system (Leerling- en Onderwijsvolgsysteem; LOVS; Cito, 2012). Actual scholastic achievement was measured twice a year by the school by means of different tests from the student tracking system.

Maths (M3). The Maths test aimed at providing an indication of the child's level of maths. The maths exercises were read aloud by the teacher, and the child indicated the correct answer on an answer sheet. The child could choose from a selection of pictures of possible answers.

Technical reading (M3). The Three Minutes Test (Drie-Minuten-Toets; DMT) aimed at providing an indication of the child's level and development of technical reading skills, and detect possible reading problems. The test consisted of three different reading cards with separate words. The child was given one minute for each card to read aloud all the words on the card.

Spelling (M3). The Spelling test aimed at providing an indication of the child's level and development of spelling skills. The test consisted of word dictations. The teacher read aloud a word in

a sentence, and the child was asked to write down the corresponding word. In order to support understanding of the word, in the first grade children were also provided with a picture of the word.

Assessment of potential for learning

AnimaLogica: test. The AnimaLogica (Stevenson et al., 2009; Stevenson, Heiser, & Resing, 2016) consisted of tasks that require inductive (analogical) reasoning (see Figure 1). The dynamic test consisted of a pre-test and a post-test of 20 items and 2 practice items. Between the pre-test and post-test two short training sessions were provided. The analogies were either provided on the computer, for the pre-test and post-test, or by means of cards (training). The analogy tasks could be solved by looking carefully into how two horizontal and vertical pictures changed. A blue lion could for example be changed into a blue camel, or a blue camel could be changed into a red camel. In order to solve an analogy, children had to come up with an inductive reasoning rule, by asking themselves which elements stayed the same, and which had changed. They had to take into account all dimensions and transformations of the items: the type of animal, its colour, the number of animals (one or two), the size (small or large), the orientation (walking in the right or left direction), and position (in the upper or lower half of the square). The post-test consisted of a parallel version of the pre-test in which the items were constructed with the same transformations, but with different animals and colours.

Animalogica: training. Training consisted of two sessions that were conducted with one week in between. Five items were trained in each session. During a training session, the test leader provided structured help by means of the ‘graduated prompts’ approach with a maximum of five help steps (e.g., Campione & Brown, 1987; Resing & Elliott, 2011). When a child could not solve an item by him or herself, step-by-step help was provided starting with general, metacognitive prompts, followed by prompts becoming increasingly more specific, and the last step being modelling. After the child had solved an item, in order to check children’s understanding, the child was asked why this was the correct solution. Regardless of whether the child’s explanation was correct, the test leader then explained why the item was solved correctly. The training intervention aimed to indicate the potential for learning of each individual child, operationalised in the number and the type of help the child needed to solve the test items. The children were categorised into instructional-needs categories. These categories were based on the number of correctly solved items in pre-test and post-test, the number and type of prompts they needed, and the progression in the number of correctly solved items from pre-test to post-test. The children’s instructional needs were evaluated in the reports drawn up for the teachers. A schematic overview of the training procedure is enclosed in Appendix A.

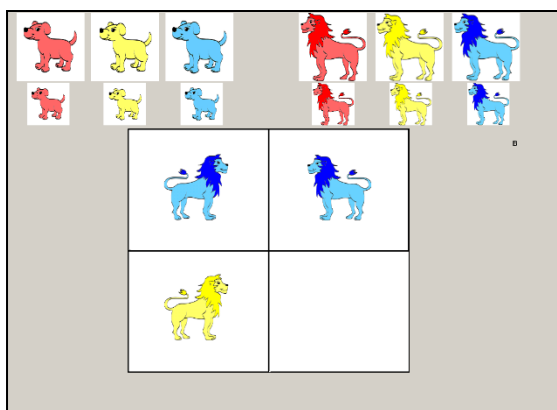


Figure 1. The computer-based version of the AnimaLogica with answering cards. The paper-based version used during the training sessions looked similar with the exception that students answered with concrete cards.

Computer mouse skills

To measure children's computer mouse skills, a short test was created for this study. The first task was a clicking task; children were asked to click on the pictures of turtles on a computer screen to make the turtle 'turn the other way' towards a pond where they could go swimming. The next task was a dragging task, where children had to drag the turtles, one by one, to the pond on the other side of the screen. The computer recorded execution time and clicking and dragging errors.

5.3 Results

Effect of the training

The analysis of variance (ANOVA) results revealed that the children who were categorised as strong functioning by their teachers had significantly higher pre-test scores ($F(1, 28) = 7.62, p = .01$) than the children who were categorised as weak functioning (see Figure 2). Then, the effect of training on the progression from pre-test to post-test was explored, and whether potential progression could differentiate between the children in the weak and strong functioning groups. A repeated measures ANOVA was conducted with Session as within-subjects factor, Condition as between-subjects factor, and the scores on the AnimaLogica as dependent variable. The results revealed that children, regardless of their condition (strong functioning, weak functioning), generally showed a significant progression in performance after training (Wilks' $\lambda = .26, F(1, 28) = 80.29, p < .001, \eta_p^2 = .74$). In contrast to our expectations, children in the weak and strong functioning groups benefited to an equal extent from dynamic training, and children in both conditions showed parallel mean progression in analogical reasoning (Wilks' $\lambda = .99, F(1, 28) = .16, p = .69$) (see Figure 2). These results indicated that the two groups differed in their initial level of reasoning, but not in their learning ability during the dynamic test.

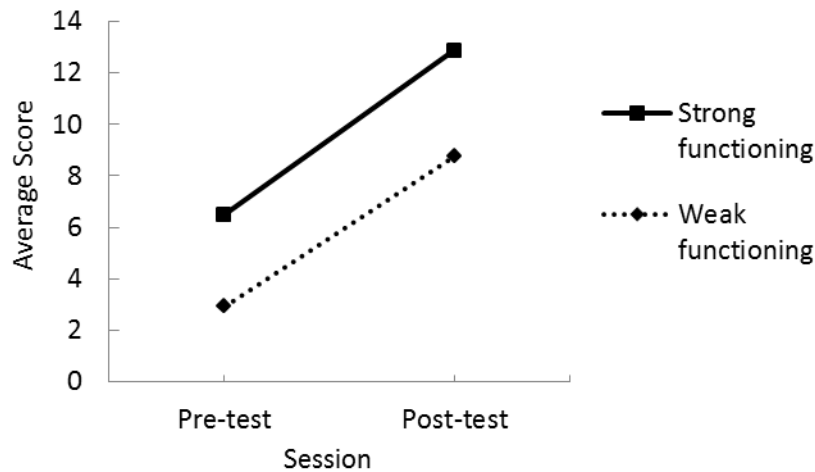


Figure 2. Number of correctly solved items per session and condition (weak functioning/strong functioning)

Individual differences in progression

The progression from pre-test to post-test (gain score) is provided for each individual child, per condition, in Figure 3. Almost all children in both conditions showed progression from pre-test to post-test after training. In general, it also becomes clear from Figure 3 that the two conditions (weak functioning/strong functioning) did not differ substantially in progression. We do, however, see substantial variation in the extent to which individual children in both conditions showed progression.

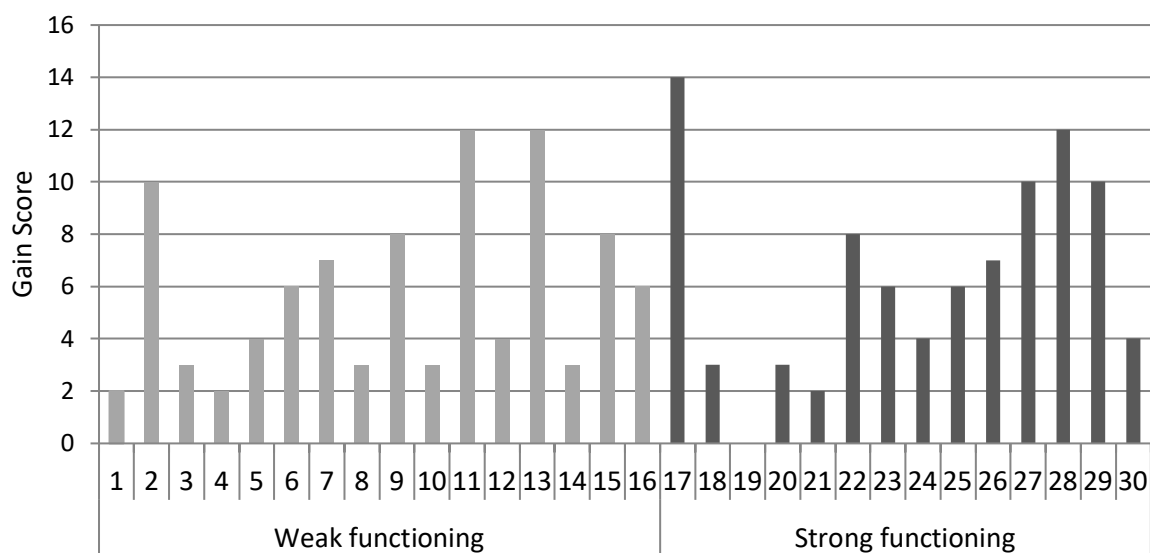


Figure 3. Progression from pre-test to post-test per condition (weak functioning/strong functioning) and per individual child

Evaluation of the training intervention

During both training sessions, the children, regardless of condition, in total needed an average of 12 (of 30) prompts; on average they needed 8 prompts for training session 1, and 4 prompts for training session 2. The total number of prompts, the metacognitive, cognitive, and modelling steps that the children in both conditions needed per session are provided in Figure 4. Both the weak and strong functioning groups showed a decrease from training session 1 to 2 in the total number of prompts. During the first training session, the children needed cognitive and modelling help steps frequently, in addition to the general metacognitive prompts. During the second training session, repeated general instruction, and metacognitive prompts were often sufficient. The children in the weak functioning group on average needed a total of 15 prompts, whereas the children in the strong functioning group needed a total of 8 prompts on average. A one-way ANOVA revealed that the difference in the number of prompts needed was significant ($F(1, 28) = 6.32, p = .02$).

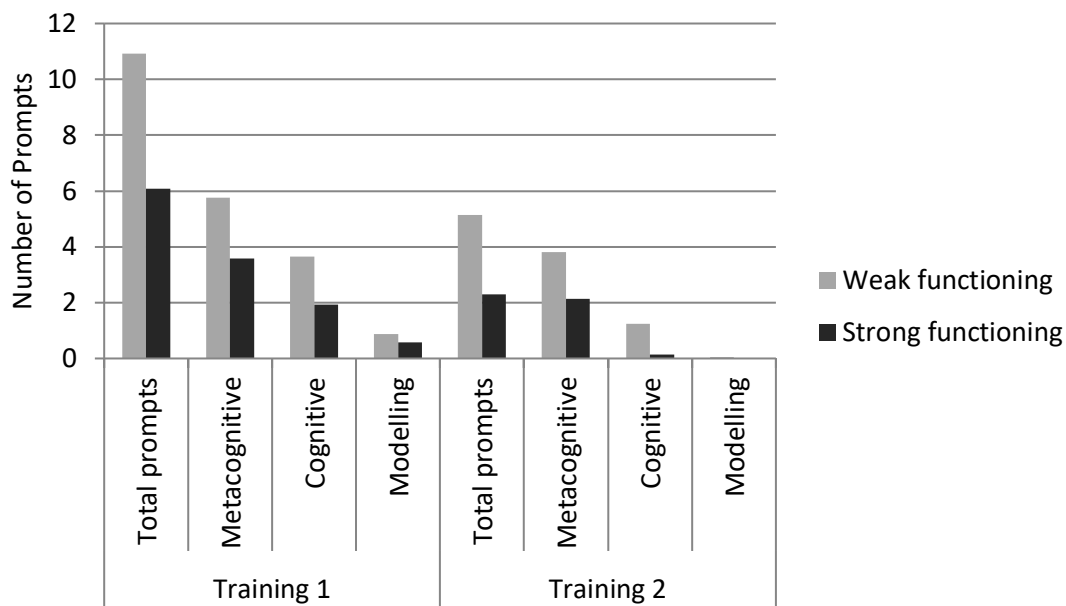


Figure 4. Total number of prompts, metacognitive, cognitive and modelling prompts/scaffolds per training session and per condition (weak functioning/strong functioning)

Individual differences in prompts needed

In Figures 5 and 6, the amount of help that was required during training session 1 and 2 is provided per child and condition, divided in the three types of prompts (metacognitive, cognitive and modelling). These figures also show that, in general, children needed more cognitive and modelling help steps in training session 1, whereas during the second training session metacognitive prompts

were sufficient more frequently. Again the figures show substantial individual variation in the number and type of prompts needed.

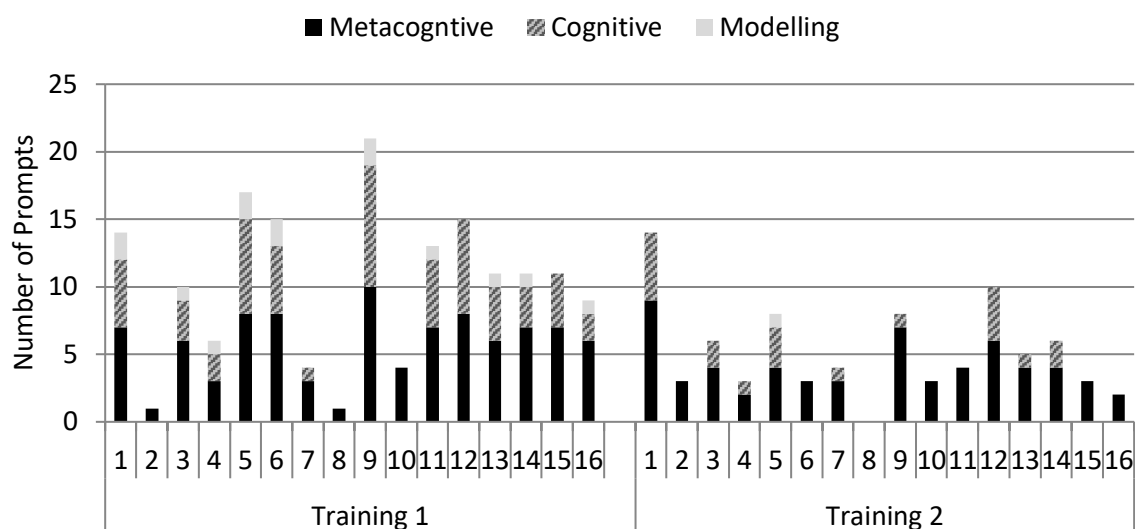


Figure 5. The number of metacognitive, cognitive and modelling prompts for the children in the weak functioning group

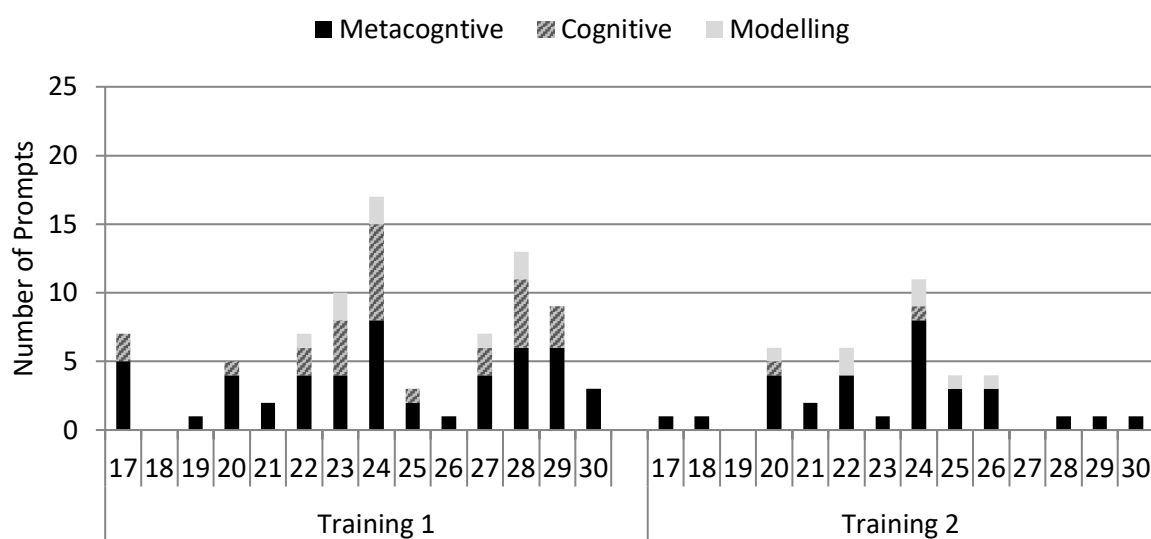


Figure 6. The number of metacognitive, cognitive and modelling prompts for the children in the strong functioning group

Relationship between dynamic test measures and scholastic achievement

In addition to the AnimaLogica test, scholastic achievement tests were conducted amongst all children. It was examined whether dynamic test results (prompts, post-test) would show a stronger relationship with scholastic achievement measures than static test results (pre-test). In Table 1, an overview is provided of the results of a correlational analysis between the Maths, Reading and Spelling test results, on the one hand, and the results of the AnimaLogica pre-test, post-test and the number

of prompts provided during training on the other hand. It was found that the Maths scores were significantly related to both the AnimaLogica pre-test, post-test and the number of prompts required during training. Pre-test and post-test results were found to have a similar relationship with the Maths results, but the relationship found between the Maths test and the number of prompts was the strongest. It was also found that the correlations between, on the one hand, the AnimaLogica pre-test, post-test and prompts, and, on the other, Reading and Spelling test scores were lower, especially as regards the total number of prompts.

We can conclude that the pre-test and post-test scores were related to a comparable extent to the scholastic achievement measures, and that only the total number of prompts required was shown to be substantially stronger related to the Maths results than both the pre-test and post-test scores. It must be taken into account that our sample was relatively small, and that the test results were not distributed normally, as children from strong and weak functioning groups were included in the present study.

Table 1. Correlations between the scholastic achievement tests and the pre-test, post-test and total number of prompts, per condition (weak functioning/strong functioning)

		Pre-test	Post-test	Total number of prompts
Maths	Total	.364*	.393*	-.510**
	Weak functioning	.074	-.085	-.124
	Strong functioning	.266	.858**	-.811**
Reading	Total	.310	.258	-.089
	Weak functioning	-.530*	-.319	.592*
	Strong functioning	.427	-.206	.028
Spelling	Total	.481**	.409*	-.253
	Weak functioning	.036	-.028	.276
	Strong functioning	.648*	.708**	-.651*

Note. * Correlation is significant at a .05 level, ** Correlation is significant at a .01 level

Therefore, we also explored whether the relationship between static and dynamic measures, and scholastic achievement tests was different when comparing children who were categorized as either weak or strong functioning by their teacher. The correlations are provided per condition in Table 1. It was found that amongst the children categorised as strong functioning, the relationship between Maths results and the scores on the dynamic post-test and the number of prompts provided during training was significant and strong, and had increased with respect to the relationship with the pre-

test scores. A similar relationship could not be found for the Reading scores, where dynamic measures were not found to be significantly related to reading achievement at all, amongst the children in the strong functioning group. Amongst the children in the weak functioning group, however, a relationship was found between both the pre-test and number of prompts provided during training, and the reading test results. However, the direction (positive/negative) of these correlations was in contrast with our expectations. The Spelling results of the children in the strong functioning group, nevertheless, showed strong correlations with all the AnimaLogica test results.

Relationship with teacher's estimation of potential for learning

Finally, the relationship between the dynamic test measures and the teacher ratings of the children's performance on eight domains, Maths, Language, Technical reading, Spelling, Writing, Reasoning, Potential for Learning, and Overall rating of scholastic achievement was examined (see Table 2). A relatively strong relationship was found between the teacher ratings of Maths, Spelling, Reasoning, and Potential for Learning, and the AnimaLogica scores. The teacher ratings of the children's writing skills were not related significantly to the AnimaLogica test measures. In contrast with our expectations, the teacher ratings as to the potential for learning and the children's scholastic achievement were found to show similar relationships with the dynamic test measures and the static pre-test scores.

Table 2. Correlations of teacher ratings and static and dynamic test measures

	Pre-test	Post-test	Total number of prompts
Maths	.441*	.415*	-.620**
Language	.414*	.365*	-.264
Reading	.480**	.361	.304
Spelling	.504**	.466**	-.423*
Writing	.202	.200	-.189
Reasoning	.611**	.512**	-.515*
Potential for learning	.471**	.510**	-.475**

Note. * Correlation is significant at a .05 level, ** Correlation is significant at a .01 level.

Evaluation of the dynamic test reports

The results of the dynamic test and the instructional needs of the children were used to draw up individual reports (see Appendix B). In these reports, the results on the pre-test were described, as well as the type and amount of prompts the child benefitted from, the extent to which the child showed progression, and the results of the post-test. Individual recommendations for the child's teacher were drawn up on the basis of these data. The individual reports were evaluated with the

teacher by the educational consultant, and the school's internal educational supervisor. These consultations were supported by practical examples, for example as to the type of questions to ask a child when giving instructions in the class room, and how to prevent deviating to a level of instructions that is too low. The evaluation sessions with the teachers and the school's internal educational supervisor showed that the recommendations from the dynamic reports provided the teachers with advantages. They explained that they had been given more insight into the type of instruction that suits an individual child, and that this is not always related to his or her level of cognitive functioning. They also noticed that children of strong and weak functioning groups could have similar instructional needs, and that children with comparable instructional needs could have a different cognitive level. In addition, it had become clear that some children could potentially show better achievements with specific instruction or support. This last finding needs to be researched more closely. Of course, the dynamic testing method could only provide limited information as regards the educational needs of the individual children. Usually, the educational consultants draw up several educational needs, as a result of assessment, in terms of what they deem necessary for that individual child. In cooperation with the school, they then evaluate which aims are feasible for the teacher, and how the action plan for the child will be operationalised. In the present study, however, the instructional needs of the children played a central role, as well as the distinction between metacognitive, cognitive and modelling prompts and a potential development in the need for these specific prompts.

The internal educational supervisor of the school in question stated that the school had obtained specific information on the basis of which children could be clustered into groups pertaining to their instructional needs. Teachers explained that they found highly insightful the extent to which children progress, from what type of instruction (cognitive prompts, metacognitive prompts, step-by-step instructions) the child benefits, and the amount of instruction an individual child needs. They also noted that these data have advantages over results of intelligence or scholastic achievement tests, as the data obtained through the latter measures are based on static tests and often do not provide any insights into specific instructional needs. The teachers indicated that the dynamic test data provided recommendations that could be applied in daily educational practice. The internal educational supervisor added 'I already have a child in mind for whom taking a dynamic test would be beneficial'.

5.4 Discussion

The main aim of the present study was to reveal the advantages of dynamic testing for education in practice. First, it was examined whether the training intervention of the dynamic test was effective. The training intervention was indeed shown to be effective for children from both conditions (weak functioning/strong functioning). The children in the weak functioning group had a lower average score on the reasoning test than children in the strong functioning group. The effect of the training

sessions, however, seemed to be similar for both groups; no difference was found in learning ability when comparing the children who were categorised as weak functioning with their strong functioning counterparts. The quantitative part of the study, in addition, showed that the children, regardless of their categorisation as strong or weak functioning benefited to very different extents from the graduated prompts training intervention. In both groups, individual differences were found: children that benefited highly or only marginally from training, and children that differed in the type and number of prompts they needed during training.

During the training interventions, a development could be found in the type of prompts necessary. During the first training session, in addition to general, metacognitive prompts, the children needed especially cognitive and modelling prompts in order to come to a correct solution. During the second training session, it was found that most of the prompts needed were metacognitive. A possible explanation of this development in instructional needs could be that the children experienced a learning effect. In the first training session, the children benefited from going through the whole-task approach, before they could master the solving process. In the second training session, most of the times the children knew how to solve the item and arrived at a correct answer independently. They seem to have learned what to focus on in solving these analogical reasoning tasks, and how they could approach these tasks independently. As a result, a great number of the children did not need any help other than metacognitive prompts, and they did not need guidance through the entire solution procedure.

As mentioned above, the children were categorised into two groups, strong and weak functioning, on the basis of the teacher ratings. In order to show the advantages of dynamic testing, it was examined whether the dynamic test scores showed a stronger relationship than static test measures with the scholastic achievement of the children, both by means of teacher ratings and in the form of actual scholastic achievement as measured with a student tracking system. The relationship between the scores of the children on the AnimaLogica and their school achievement scores were examined. The results lead to the conclusion that the static and dynamic test scores of the children were related significantly to the Maths and Spelling test results. This relationship was particularly strong amongst the children categorised as strong functioning. The dynamic measure, the number of prompts that the children needed during training, had the strongest relationship with Maths test scores. The children's analogical reasoning ability was found to be related only to the Reading test results of the children in the weak functioning group. The actual scholastic achievement results, as measured by a student tracking system, was overall found to be related to an equal extent to the dynamic test measures and the static pre-test scores.

The relationship between, on the one hand, the teacher ratings as to the children's potential for learning and their scholastic achievement, and, on the other, the dynamic measures and the static

pre-test scores was also examined. The teacher ratings as to the children's achievement as regards Maths, Spelling, Reasoning and Potential for Learning were found to be strongly related to the static and dynamic test measures of the AnimaLogica. The teacher ratings were found to be related to an equal extent to the dynamic test measures and the static pre-test scores of the AnimaLogica. It seems that the teacher rating as to the children's cognitive functioning in several domains and the children's general learning ability is primarily based on the obtained scholastic achievement results and to a lesser extent to instructional needs, learning strategies and other individual characteristics regarding learning.

The present study aimed to make the advantages of a dynamic test visible. This form of testing is as yet applied only incidentally in practice due to the extra time it takes to conduct these tests, in spite of the fact that the testing time of the AnimaLogica was limited as pre-test and post-test were conducted on the computer. In principle, this allows for testing several children at the same time. In addition, we would argue that the procedure allows us to examine the learning processes of the children in great detail, including the progress in performance and the types of prompts the children need. The evaluation consultations with the teachers and internal educational supervisor, have shown that especially the practical recommendations obtained through the dynamic test reports is highly valued. First of all, the dynamic test results helped the teachers gain insight into the children's potential for better performance as a result of different instruction/support. The second result of the dynamic test is related to this and concerns the operationalisation of this instruction and support. This is the part of learning that cannot be tapped into by static tests, and the part of a dynamic test that, according to the teachers and internal educational supervisor cooperating in this study, is of added value to education in practice. The teacher will obtain more insight into instructions suitable for individual children, and will find out that this may not always be related to his or her level of intelligence. The different types of instructional needs can be found amongst all levels of intelligence. A child categorised as weak functioning by a teacher could have similar instructional needs as a child categorised as strong functioning. Both children could, for example, have a need for the metacognitive prompt 'look first, then think, and then act', which will allow both to progress substantially in performance. If the instructional needs of children within a classroom are measured by means of a dynamic test, children would be provided with the instructions that are tailored to their needs (e.g., Bosma, 2011; Bosma & Resing, 2008, 2012). Of course, further research, particularly in the domains of knowledge tested in school, is necessary to examine this assumption in more detail.

Adapting more to the instructional needs of individual children, and, in this way, tailoring more to the individual child's educational needs, seems to increase the chances that the children are instructed within their zone of proximal development (Vygotsky, 1962, 1978). The dynamic test would, in line with this theory, be a better predictor than static tests of scholastic achievement (Beckmann,

2014), but the present – pilot – study has not found such a relationship. A larger sample with a control group, including a long-term training intervention, dynamic scholastic achievement measures, and a follow-up study should provide more clarity in these respects. The advantages of a dynamic test are, however, not just limited to this predictive value, but mainly in the applicability of practical recommendations resulting from the training intervention (Elliott, 2003; Lauchlan & Carrigan, 2013; Lauchlan & Elliott, 2001). This is supported by the positive evaluations of the teachers and internal educational supervisor cooperating in the present study.

Appendix A

Schematic overview of the graduated prompts approach

Type of prompt	Verbal instruction
Metacognitive	Examiner asks whether the child remembers what he or she needs to do. Examiner explains that the child needs to look at the puzzle and think carefully. The child is then told that he or she can solve the puzzle by placing the correct animal(s) in the empty box.
Metacognitive	Examiner explains changes from A to B and tells the child that C needs to be changed in the same way.
Cognitive	Examiner guides the child through the problem-solving steps from A to B and A to C.
Cognitive	Examiner summarizes all horizontal and vertical changes.
Modelling	Examiner gives step-by-step modelling scaffolds to solve the problem.

Appendix B

Example of a dynamic test report

In this report, you will find an overview of the results of AnimaLogica, a reasoning test conducted at your school by educational consultants of the educational advisory service. The results of the teacher rating questionnaire that you have filled out, are also incorporated into this report. This study is still ongoing, and the results should be considered as a first estimation of the learning achievements and ability of the child. These need to be interpreted with care. The results have limited validity, as the test used is yet to prove its applicability in educational practice.

Personal details

Name : X
Age : 6 years, 9 months
School : Elementary school
Period of testing : April-May 2015

Observations

X was tested four separate sessions. The average duration was 20 minutes per test session. X entered the room as a cheerful and spontaneous girl, who was motivated and talkative. X sometimes got distracted during the task. Her attention would then have to be redirected to the task.

Teacher rating

The teacher has categorised X as follows on a 5-point scale (1 equals the 20% lowest, and 5 the 20% highest performing children).

	Maths	Language	Technical Reading	Spelling	Writing	Reasoning
Score	2	3	2	2	3	3

Scholastic achievement

The overview below provides the Cito school results of X.

	Maths E3	Spelling M3	Technical reading E3
Score	E	E	C

Learning ability

The teacher has categorised X's learning ability on a 5-point scale (1 equals the 20% lowest, and 5 the 20% highest performing children). Both X's potential for learning and the overall estimation of her scholastic achievement were rated a 2.

Results

AnimaLogica, a dynamic reasoning test, was conducted as part of the study.

Pre-test

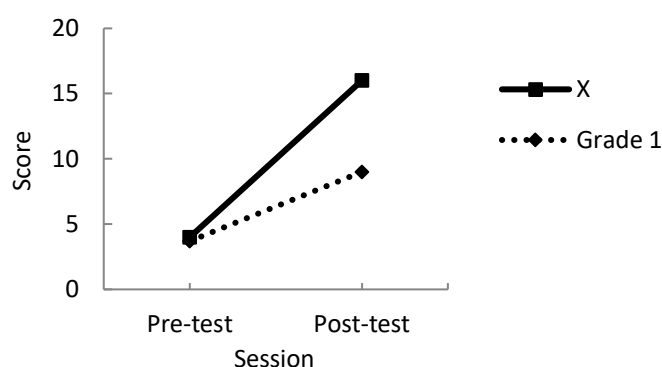
X could solve 4 of 20 items correctly in the pre-test. This is an average score compared with other children from grade 1 who participated in the study. X mainly used partial solution strategies. X seemed to understand the principle of analogical reasoning, but had difficulty paying attention to all aspects (the different transformations) of a task. She seemingly did not see some transformations, such as the animal's position.

Post-test

In the post-test, X could solve 16 of 20 items. Her result is above average as compared with other children from grade 1 who participated in the study. Now, X mainly used correct solution strategies, which led to correct answers. She has seemingly mastered analogical reasoning. X took all possible transformations into account, and took more time checking her answer.

Training

X was provided with a training intervention between the pre-test and post-test. She needed an equal number of prompts as her age-mates in order to come to a correct answer. In the beginning, X needed a clear explanation. Then, instructing her that her answer was not yet correct, and that she had to think carefully first before giving her answer seemed to suffice. A graphic overview of X's progression compared to her age-mates is provided in the figure below.



Conclusion

Without adapted instruction, X performed on an average level. After training and adapted instruction, her performance increased to an above-average level. X needed an average amount of help in order to arrive at correct answers. X's progression as a result of training was above average. The training has shown that a general instruction is not sufficient for her. An extended instruction that included going through the steps and a practice opportunity, enabled her swiftly to solve items correctly. Then, it sufficed to tell her that she had to think carefully first, before giving an answer. The results show that X understands the principles of analogical reasoning, and is capable of explaining these to others.

Recommendations

On the basis of the results above, the following is advised: X benefits from an elaborate first instruction with a lot of specific help in making assignments. This is help in which an activity or procedure is explained step-by-step, and is demonstrated precisely. After a short time, general instructions are sufficient for X. The instruction: "Do you know what you have to do? First have a good look, then think carefully, and then give an answer" works well for X. In addition, it is beneficial to tell X to check her answers.

