CHAPTER 3

Need for Instruction:

Dynamic Testing in Special Education

The contents of this chapter are in submission:

Tirza Bosma & Wilma C. M. Resing.
Abstract

The aim of this study was to examine the contribution of dynamic testing in the measuring of children’s need for instruction and to explore responses of special education teachers to dynamic testing results. Thirty-six 10 to 12-year-olds with a moderate to mild intellectual disability and their teachers participated. Children in the experimental condition were dynamically tested. This test included a graduated prompts training and an analogy construction task; children in the control group were administered a static pretest and posttest only. Outcomes of the assessments were reported to teachers. Teacher-child interactions were observed twice, learning potential ratings were gathered and teachers were interviewed regarding reports. The results showed that dynamically tested children achieved significant higher posttest scores than untrained children. Dynamically tested children, even with comparable low IQ scores, varied in their need for instruction, measured by the number and type of prompts they required during training as well as during construction problems. Teachers appreciated the reported dynamic testing outcomes, although substantial changes in teaching practice were not observed. Supplementing intelligence testing with a dynamic testing procedure is recommended to obtain a more accurate description of children’s educational needs.

Acknowledgements

We thank M. Rodenburg for helping us to collect and code the data.
Introduction

Intellectually disabled children, identified on at least an IQ below 70, are often characterized as being inefficient learners needing repetitive instructions and practice. These children are frequently allocated to special education in small groups, consisting of children with comparable intellectual disability levels. The underlying assumption is that children within such an intellectual disability category presumably have relatively equal and homogeneous learning characteristics and therefore will be able to benefit from equal forms of support and instruction. However, there appears limited evidence and support for instruction and service delivery based on homogeneous categories of learners and diagnostic classifications (Caffrey & Fuchs, 2007; Ysseldyke, 1987). The main reason might be that children with intellectual disabilities are not such a homogeneous group when we look at their potentials for learning. Groups are mainly composed based on the outcomes of intelligence testing, which can be seen as indicators of existing required knowledge, meaning the end products in the form of test scores rather than the process of learning. Various authors stated that standard intelligence test results hardly provide information regarding the process of learning and therefore fail to contribute to better instructional planning (e.g., Haywood & Lidz, 2007; Ysseldyke, 1987). This was stated in particular for children with intellectual disabilities (e.g., Feuerstein, Rand, Hoffmann & Hoffmann, 1979; Taylor, 1987), among other reasons, because of the often found floor effects on intelligence tests for this group (e.g., Hessels-Schlatter, 2002a).

Intelligence test outcomes of children with intellectual disabilities in most cases provide only very little information about the why and how of the low scores, and their IQ’s may therefore not correctly reflect the student’s potential for learning (e.g., Carlson & Wiedl, 2000). Within the group of children with intellectual disabilities, in fact, children may very differentially profit from instruction (Reschly, 1987) and need a variety of types, amounts and even levels of instruction to be able to show learning
progress. Caffrey and Fuchs (2007), in their review of learning differences between children with learning disabilities versus mild intellectual disabilities, concluded that children with learning disabilities learned more quickly and profited more from instruction. They remained skeptical however about this research finding for various reasons. The number of available studies was limited, studies included had unclear instructional contexts; there were questions regarding the validity of the classifications used in these studies, and finally the authors had the assumption that overlap existed between the two categories of children regarding their type of learning. Hence, they plead for more research on theoretical foundations and effectiveness of instructions.

Research on dynamic testing might be one way to address the need for studies regarding the relation between educational needs and individual learning variability. Outcomes of dynamic testing aim to inform, among others, about the child’s educational needs (Campione & Brown, 1987), by exploration of the nature and amount of help, assistance, and instruction a child needs to solve cognitive-intellectual problems and school achievement tasks. Furthermore, it is assumed that outcomes of dynamic testing may inform recommendations for teaching practices (Grigorenko, 2009; Haywood & Lidz, 2007). The main purposes of our study therefore were first to investigate the potential for learning of students with intellectual disabilities, in particular the type and amount of instructions they were profiting from, while dynamically tested, in order to guide classroom instructions. Secondly, we explored how special education teachers respond to dynamic testing results in their instructional practice and their opinions regarding the usefulness of the dynamic testing results for classroom instructions and educational planning.

Dynamic assessment procedures have been developed as alternatives or supplemental measures to intelligence tests. Standard assessment, such as intelligence testing, is administered without giving any feedback in the testing process, while in dynamic testing the response of the child elicited to prompts and feedback is recorded when solving cognitive tasks, usually complex in nature. Several studies have shown that measures for children’s learning potential considerably vary while their IQ-range is much
smaller (e.g., Resing, 1997, 2000), even within the (very) low range (Bosma & Resing, 2006; Hessels-Schlatter, 2002a).

Studies concerning the learning potential of mild intellectually disabled children have often been carried out to show possible underestimations of children’s potential when using intelligence tests (e.g., Feuerstein, Rand, Hoffmann & Hoffmann, 1979; Tzuriel & Klein, 1985). Schlatter and Büchel (2000) measured the learning capacity of adolescents with severe intellectual disabilities. The intervention phase of their dynamic ARLT test consisted of a scripted protocol with hints and feedback for solving perceptual analogies. They reported different categories of learners: gainers, non-gainers, and an undetermined group. Hessels-Schlatter (2002a) showed that students categorized with the ARLT as “gainers”, having higher potentials for learning, despite their low IQ’s, benefited from a cognitive training program, while “non-gainers” did not profit. Results of a study with the ARLT administered to Dutch kindergarten children with mild intellectual disabilities showed variation in potential for learning across the reported IQ’s as well. Besides, additional measures derived from dynamic testing provided extra information about the children’s understanding and application of analogical reasoning (Bosma & Resing, 2006). Despite these valuable results regarding individual differences in learning potential, the ARLT-training procedure has been set up to distinguish children who profit from cognitive programs and more demanding schooling from children for whom these programs or education are less appropriate, and the training procedure is not specifically designed to guide education tailored to the needs of the individual student.

A structured approach within dynamic assessment that focuses in particular on the need for instruction is the graduated prompts technique (e.g. Campione & Brown, 1987; Resing, 1997, 2000). In recent studies on typically developing children, including children with a different cultural background, Resing, de Jong, Bosma, & Tunteler (2009) showed that such a dynamic testing procedure with a highly structured training, provides information about potential for learning of children, in particular the minimum number of prompts children needed and their variability in the number of meta-cognitive and cognitive prompts required to solve the tasks independently). The training
procedures that were used not only focused on teaching specific problem solving strategies, but also included instructions regarding metacognitive skills, such as planning and checking (Campione, Brown & Ferrara, 1982; Resing, 2000). In addition to information regarding the needs for instruction, an extra phase in the dynamic testing procedure, in which learned principles have to be applied, may provide information to guide educational interventions. Young children with intellectual disabilities were, for example, able to construct analogy problems on a reversal task after a dynamic test procedure (Bosma & Resing, 2006).

Although the description of the needs of children is important and a first step to guide educational interventions, the potential for learning of children may only start to unfold in an environment that both focuses on the potential of the child and provides opportunities to accommodate his or her needs (Elliott, 2003; Hessels-Schlatter, 2002a). Several aspects are important in unfolding the potential for learning, one is teacher expectation. Feuerstein and colleagues (1979) pointed out that assignment to special education classes may act as a self-fulfilling process: teachers have low expectations of students and act accordingly to that premise. Incidentally, research findings showed that outcomes of dynamic assessment may lead to different teacher expectations regarding a child’s learning ability (Delclos, Burns, & Kulewicz, 1987) and more realistic teacher estimations regarding a child’s potential for learning potential (Bosma & Resing, 2010). Another aspect is the estimated value of dynamic testing results by teachers, in particular regarding the applicability of the assessment outcomes in their teaching practice. In previous studies it was found that teachers in regular education valued dynamic testing results as useful for guiding their instructions for individual children with specific math needs, but not directly for the typically developing children in their classes (Bosma & Resing, 2008; 2010). Results were also considered as meaningful and useful for planning educational interventions regarding students with math difficulties (Bosma & Resing, 2010), and typical developing preschool children (Hulburt, 1995). However, changes in instructional practices were hardly visible or not acknowledged by teachers, which might indicate that teachers did not (yet) implement the provided recommendations in their practice.
The aim of the current study was to investigate the contribution of dynamic testing, including a graduated prompts training, in measuring the need of instruction of children with intellectual disabilities, compared to their static intelligence test results. A second aim was to explore how special education teachers respond to dynamic testing results, especially regarding changes in their instructional practice, their ratings of learning potential and to explore their opinions regarding the practical value of the results for their classroom teaching. It was assumed that studying 10-12 year old children identified as intellectually disabled would provide us additional information regarding their specific instructional needs and would give us an impression of the use of dynamic testing outcomes in special education practices.

Based on studies of Resing (1997, 2000) and Resing et al. (2009) we firstly expected that children who received a graduated prompted training as part of the dynamic test, would achieve higher and more differentiated scores on the posttest compared to the non-trained children in the control group. Secondly, it was expected that dynamically tested children would show variation in their need for instruction during training, defined as the total number of prompts, and more specifically the number of metacognitive, general cognitive, specific cognitive prompts or the full explanations required. It was also expected that these variations do not necessarily relate to their IQ. Such variability was found in earlier research in both typical developing children and children with math learning difficulties (Resing et al., 2009; Bosma & Resing, 2010). Thirdly, we expected a strong relation between the posttest score and the scores on the analogy construction task as well as variability in the need for assistance in constructing these.

Regarding the responses of special education teachers towards the dynamic assessment results we explored whether changes in teacher-child interactions could be observed after providing testing results to teachers and whether teachers changed their ratings of children’s learning potential in response to the reported dynamic testing results. In addition, teachers’ opinions regarding recommendations based on dynamic testing were explored. In particular the practical use of the dynamic testing results and recommendations was inquired.
Method

Participants

Participants were 21 boys and 15 girls all identified as having an intellectual disability (Mean IQ = 58.6; SD = 5.6). All children attended special education classes in two schools for students with intellectual disability in middle size cities in the Netherlands. Mean age of the children was 11 years and 6 months, ranging from 10.4 to 12.5 years. One third of the children were diagnosed as having autism spectrum disorders, ADHD or Down’s syndrome as co-morbidity. Informed consent was obtained from parents for test administration and access to school records. Teaching experience of the four teachers of the participating children ranged from 4 to 20 years.

Design

The study employed a pretest-posttest-control group-design with randomized blocking based on recent results of intelligence measures, retrieved from children’s school records. The experimental group was administered a pretest, training, posttest and a construction task of an adapted version of a dynamic test (CCPAM, Tzuriel & Galinka 2000). To children in the control group only the pretest and posttest were administered. All teachers had experimental-group as well as control-group pupils in their classroom. Before and after test administration interviews with teachers and observations were conducted. Table 3.1 gives an overview of the design of the study.

<table>
<thead>
<tr>
<th>Table 3.1. Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
</tr>
<tr>
<td>--------------------</td>
</tr>
<tr>
<td>Experimental</td>
</tr>
<tr>
<td>Control</td>
</tr>
</tbody>
</table>
**Instruments and Dynamic Training Procedure**

*Children’s Conceptual and Perceptual Analogies Modifiability test (CCPAM; Tzuriel & Galinka, 2000)* is an existing dynamic test in the domain of analogical reasoning. The test includes both conceptual and perceptual analogies. In our study we used only the version with conceptual analogical problems, which, according to Lifshitz, Tzuriel and Weiss (2005) was shown to be of lower complexity level for people with intellectual disabilities than the perceptual analogies. Tzuriel and Galinka (2000) reported internal consistencies for conceptual pretest and posttest of $\alpha = .73$, and $\alpha = .84$, respectively. The test we used consisted of a pretest, a training session, and a posttest, each incorporating 20 classic matrixes in the A:B::C to D1-D4) format.

The adapted CCPAM-training. The aim of the original training of the dynamic test, as described in the manual (Tzuriel & Galinka, 2000), is to teach a child the dimensions for analogical reasoning, understanding analogical principals, a systematic task approach and to improve the efficiency. However, the manual does not provide specific training instructions or specified series of prompts with a fixed order. To be able to capture children’s potential for learning and particularly their need for instruction measured by the number and type of prompts, we constructed an adapted training procedure, equaling the structured graduated prompts protocol of the Learning Test for Inductive Reasoning (LIR; Resing 1990, 1993, 2000). The training method we used for the current study consisted of a standardized series of seven hierarchically structured prompts; which order was based on analyses of the solving process of analogies (e.g., Resing, 1990; Sternberg, 1985). Prompts were given if a child could not independently solve a problem or made mistakes, and started with meta-cognitive prompts for example “Have you done this before? Does that help you to find the answer?”, followed by more concrete and task specific ones. A more specific cognitive hint, for example is: “Why, do you think, A and B go together? What will go with C?”; in which A, B and C stand for the illustrated objects of an analogy. In the seventh, most specific prompt, the examiner completely explained to the child what to do. For every training item the
number of prompts given (0-7) was recorded. Measures for learning potential were defined as the test score after training, the minimum number of prompts split into metacognitive, general cognitive and cognitive specific prompts, a child needed to further solve the analogical reasoning problems independently.

**Construction Analogies.** In construction analogies, an additional 8-item subtest of the CCPAM (Tzuriel & Galinka, 2000), the child had to create an analogy out of 6 pictures, with comparable analogy types as practiced in the training. The construction version of the CCPAM aims to measure transfer of the taught principles to a new but similar task. We followed here the original manual of the CCPAM, in which a series of five prompts based on graduated prompting technique was specified (Tzuriel & Galinka, 2000). These series of prompts consist of, in our terminology, cognitive and very specific visual prompts. The series starts with nonverbally providing the A:B parts to the child, followed by explaining the A:B relation, providing and or explaining the third picture and finally the whole analogy was given and explained by the examiner. The maximum score of 8 x 6 can be reached. A high score means that no or little help was required.

**Intelligence Reports.** Psychological reports were retrieved from the school records and were not older than 6 months. Summaries of each child’s diagnostic reports were discussed with teachers, including information regarding cognitive functions, learning, IQ measures, discrepancies if existed, conclusions regarding learning difficulties, task approach and on task behaviour.

**Interviews.** Teachers were systematically interviewed twice for about 10 minutes. The first interview explored teachers’ practical experience regarding adapting their teaching to individual needs, regarding individual instructional plans and regarding recommendations they considered valuable. The second interview took place after teachers received the assessment results, to evaluate the practical use of the dynamic or static assessment results and observations. In addition, in both interviews teachers rated the learning potential of each participating child on a 6 point scale.

**Classroom Observations.** Lidz’ (1991, 2003) Mediated Learning Experience (MLE) rating scale was used to capture teacher–child interactions during instruction. Teacher-child interactions were rated on the following 4-point scales: Intentionality (involvement of
teacher), Task Regulation (type of task instruction), Praise and Feedback (frequency of positive feedback), Challenge (challenging children’s zone of proximal development, or ZPD) and Informing Change (informing the child’s achievement).

**Dynamic Reports.** Results of the dynamic testing were discussed for each child with teachers after the assessment. All reporting consisted of a short description of the experiment and outcomes of dynamic testing. The reports further included learning potential measures as rated by the teacher, observations, pre- and post test results, error-analysis and task approach, amount and type of help needed during training and the results of the construction task. For the control group the results of the classroom observations, the learning potential rating and the pretest and posttest scores, including error analysis, were reported.

**Procedure and scoring**
Administration of pretests and posttests took approximately 15 minutes per child per session; duration of the training and construction session varied from 10 to 30 minutes. Prior to testing the examiner reported summaries of the static intelligence reports of all participating children to the teachers, teachers were interviewed and classroom observations were conducted. After assessment sessions (see Table 1) the examiner reported the outcomes to the teachers. Three weeks after the briefing of the test results the second interview with teachers was conducted.

To answer the various research questions in our analyses children in the experimental group were categorized in groups based on IQ-scores, pretest scores, posttest scores or number of prompts. We used median scores to divide children in groups. Based on a Median score of 58, a low IQ-group ($N = 9$; IQ between 48 below 58) and a moderate IQ-group ($N = 8$; IQ between 59 and 70) were created. Pretest and posttest categories were each based on below average ($\leq 8$) versus above average ($\geq 9$) scores and categories of relative few or relative many prompts were based on a median score of 33 total prompts.
**Results**

Prior to answer the research questions it was analyzed whether experimental and control group did not differ regarding age, mean IQ and pretest scores by conducting a one-way MANOVA with condition as independent variable and age, IQ and pretest scores as dependent variables. In Table 3.2 the means and standard deviations are shown for both conditions on each dependent variable, including the posttest scores. As expected, no differences were found ($p = .80$). In addition no difference was found in regards to other diagnosis of the child, such as ADHD.

To test the hypothesis that children, who received a graduated prompt training as part of the dynamic test, would achieve higher scores on the posttest compared to children in the control group, who did not receive any training, a repeated measures (RM) ANOVA was conducted. Total correct score on CCPAM pre or posttest items specified as within-subject variable was measured in two conditions (experimental versus control) across two measurement times (pretest- posttest). Of special interest was the Time x Condition interaction because if significant this would reflect the effect of the training. Outcomes revealed, as expected, a significant Time x Condition interaction: $\text{Wilks’s } \Lambda = .88$, $F(1, 33) = 4.59$, $p = .040$, partial $\eta^2 = .12$, indicating that the scores of the children in both conditions changed differently over time, as can be seen in Figure 3.1. Trained children correctly solved significantly more analogy problems after training, whereas controls did not.

**Table 3.2**. Means and standard deviations for age, IQ pre- and posttest scores

<table>
<thead>
<tr>
<th></th>
<th>$N$</th>
<th>age</th>
<th>IQ</th>
<th>Pretest</th>
<th>posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Experimental</td>
<td>17</td>
<td>138.76</td>
<td>58.41</td>
<td>8.76</td>
<td>2.81</td>
</tr>
<tr>
<td>Control</td>
<td>18</td>
<td>137.33</td>
<td>58.67</td>
<td>8.17</td>
<td>2.72</td>
</tr>
</tbody>
</table>
Inspection of the pretest scores showed that the standard deviations of the pretest scores for both conditions are moderate. For the posttest standard deviations are somewhat higher, in particular for the experimental group, indicating that not all children equally improved their posttest performance, as is visible in Figure 3.2 and 3.3; Experimental group children showed more variability in their posttest scores, whereas control group children did not. Secondly, we explored whether dynamically tested children varied in their need for instruction, defined as the number of prompts given. The number of prompts children needed ranged from 15 to 68, with a mean of 32. Seventy percent of the children needed prompts on less than 10 out of 20 items, whereas the others needed prompts on 11 through 15 items. The types of prompts children needed varied. Items on which children needed prompts, were solved in 40.6% with only metacognitive prompts, in 23.1% with additional general cognitive prompts, in 25.6% with additional more specific cognitive prompts and in 10.6% of the items children needed in addition to the whole series of prompts, a full explanation of the correct solution.
In Table 3.3 the types of prompts children needed are specified according to the total number of items solved with prompts. Only one child needed prompts on 15 or more items, in particular metacognitive prompts. Figure 3.4 shows for each child the number of items with the type of prompts each child received, ordered on the number of items with meta-cognitive prompts. These results show that children considerably vary in the number of items for which they needed help and in the type of help.

Table 3.3. Types of prompts needed during training, in percentages, for experimental-group children

<table>
<thead>
<tr>
<th>Type of prompts</th>
<th>Total number of prompted items</th>
<th>N</th>
<th>Meta-Cognitive</th>
<th>Cognitive general</th>
<th>Cognitive Specific</th>
<th>Cognitive Full Explanation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-9</td>
<td>12</td>
<td>(2.83)</td>
<td>(1.92)</td>
<td>(2.08)</td>
<td>(.83)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td>10-14</td>
<td>4</td>
<td>(5.25)</td>
<td>(2.75)</td>
<td>(3.75)</td>
<td>(1.5)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>(10)</td>
<td>(3)</td>
<td>(1)</td>
<td>(1)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>40.6%</td>
<td>23.1%</td>
<td>25.6%</td>
<td>10.6%</td>
<td>(100%)</td>
<td></td>
</tr>
</tbody>
</table>
Next, we investigated whether groups of dynamically tested children with comparable pretest scores varied in their need for instruction. Crosstabs were analyzed to compare the categories of children with a below average versus above average pretest score to the number of prompts they needed (few/many) during training. Inspection of crosstabs showed that one third of the children in the below average pretest category accordingly needed relatively many prompts, while two third needed relatively few prompts. Of children with above average pretest scores almost two third also needed few prompts, against one third who needed many prompts. The differences between the pretest categories regarding the number of prompts were as expected not significant ($\chi^2 = .032$, $p = .86$); The pretest categories did not differ in IQ, $F(1, 15) = 2.62, p = .126$. The results show that, even in a group of children with intellectual disabilities, children with comparable pretest scores vary in their need for instruction while solving analogies; categorizing children based on a static (pre) test did only correctly inform about the need for instruction in about half the cases.

![Figure 3.4](image)

**Figure 3.4.** Number of training items solved, per child, with the different types of prompts needed
Further, we analyzed the scores on the construction task. The mean score was 40.35 ($SD = 4.2$; range 29-47). All children showed independent constructing of at least one item ($M = 3.24$; $SD = 1.64$; range 1-7) out of eight items. Figure 3.5 shows the variation in the number of construction items solved independently, combined with the type of prompts needed. Inspection of the figure shows that if children needed help, the majority (58.8%) profited from a specific cognitive prompt in which the A:B relation was given, with or without verbal explanation; 29.4% of them needed even more specific prompts, and 11.8% of them needed in addition a full explanation of the solution.

Figure 3.5 Number of construction items solved with help, per child, with the different types of prompts needed (see legend)

To answer our third expectation, we examined the relations between posttest scores, number and type of prompts, duration of training, and scores on the construction task. Training time correlated considerably with the number of prompts
needed during training ($r = .61, p = .01$) and the posttest score ($r = -.47, p = .05$). The number of prompts needed correlated considerably with the posttest score ($r = -.55, p = .03$) as well, but, as shown before, this correlation is not perfect: about one third of the children achieved a high posttest score and needed many prompts, or achieved a low posttest score, after needing relative few prompts during training.

Correlations between learning potential measures and IQ were on the other hand lower and not significant. IQ correlated low ($r = .14$) with pretest scores and moderately ($r = .44$) with the posttest scores for the experimental group. For the control group, IQ correlated moderately with the pretest scores ($r = .42$) and low with the posttest scores ($r = .07$). Further analysis regarding the experimental group showed a moderate relation between IQ and number of prompts needed during training ($r = -.44$), low correlations with both metacognitive and cognitive prompts ($r = -.19$, and $r = -.32$, respectively).

We also examined whether children in the experimental group with comparable IQ scores varied in pretest scores and their learning potential scores, and analyzed differences between a moderate and low IQ-group. Results of Mann Whitney tests with IQ-group (low/moderate) as independent variable and the needed number of total prompts, meta-cognitive prompts, cognitive prompts, the scores on the pretest and posttest and the constructing task as dependent variables, showed as expected, no significant differences between the low and moderate IQ-group. Nor were differences found between the two IQ-groups for the pretest and posttest scores in the control group.

Further, a rather high correlation was found between the posttest score and the score on the construction task ($r = .62, p < .01$), indicating that at least part of the children with a higher score on the posttest could apply their knowledge of how to solve and to make analogies. Inspection of crosstabs with low and high posttest scores and the number of items children solved independently showed that 55.5% of the children with low posttest scores could construct only 1 or 2 analogy tasks (out of 8 total), while 50% of children with higher posttest scores could construct at least 3 items and 37.5% could construct 4 or more items without any help. The relation between the number of...
Chapter 3

prompts children needed during training and the score on the reversal task was moderate ($r = -.34$, n.s.). Crosstabs revealed that 44.4% of the children, who needed relative few prompts, solved only 1 or 2 items, one third 3-4 items and 22.2% was able to construct more than 5 items without help. On the contrary, more than two-third of children needing many prompts during training constructed 3-5 items without help. These results show that at least a number of children who needed relative many prompts during training were still able to construct at least some analogy problems independently.

The second part of our study focused on teachers’ responses regarding the assessment outcomes. First, possible changes in teacher-child interactions as consequence of dynamic testing were explored. Scores on Mediated Learning Experience rating scales before training for both conditions fell mostly between 1 and 2 for Intentionality, Task regulation and Praise, indicating that observed interactions could be interpreted as being intentional, task regulating and positive supporting. Challenging was observed as low; interactions that could be identified as challenging children’s zone of proximal development were observed incidentally. Interactions in which children were informed about their learning progress (subscale Change) were hardly or not at all present before testing. Repeated measures ANOVA’s on each of the subscales specified as within variables in two conditions (experimental versus control) across two measurement times (observation 1 and 2) revealed that the observed teacher-child interactions did not differentially change as a consequence of dynamic testing. No significant interaction effects for condition with any of the observation scales were found.

Next, we explored the learning potential ratings by the teachers prior and after dynamic testing. Before the experiment started teachers rated their students’ learning potential, compared to peers with intellectual disability, as above average for both conditions ($M_{\text{exp}} = 3.76; \ SD = 1.30$ versus $M_{\text{con}} = 4.11; \ SD = 1.32$). To investigate whether the ratings changed differently over time for experimental versus control-group children, a repeated measures ANOVA with learning potential ratings specified as within-subject variables measured in two conditions (experimental versus control)
across two measurement times (prior- post testing) revealed a significant declining main
effect for rating of learning potential ($F(1, 33) = 7.96, \ p = .008$), but no significant
interaction effect. Teachers’ second ratings of learning potential did relate considerably
to the number of prompts needed ($r = -.61$), but not significantly to the posttest scores
and scores on the construction task. The first and second rating of learning potential for
the experimental group related strongly to IQ $r = .62$ versus $.71$, for the control group
these correlations were also considerable: $r = .61$ versus $.441$.

To further explore teachers’ evaluations of dynamic testing results we analyzed
the interviews held with each teacher before and after dynamic testing. Exploration of
responses on the first interview revealed that teachers reported to prefer practical and
concrete recommendations, which they thought to be lacking in most of the usual
standard reports. Teachers also explained that recommendations in standard reports
were not always implemented, because these were considered as not immediately
realizable, as already practiced interventions or as known information. In the second
interview teachers evaluated the dynamic testing outcomes as new, valuable and more
useful information compared to the former static reports. Recommendations were
found more concrete and practical. Teachers appreciated the provided dynamic testing
information because of the elaboration beyond intelligence scores and deficits of the
children. One teacher noted that the idea of learning potential, in particular the measure
“number of help and instruction needed” was easier to explain to parents and “potential
for learning” could be considered as an easier starting point to talk about a child’s
learning goals.

**Discussion**

The current study showed that dynamic testing with a graduated prompts protocol
provided additional information about the specific needs for instruction of children with
moderate to mild intellectual disabilities in special education. Trained children achieved,
as expected, higher scores on the posttest and showed more variation in their scores
than untrained children. These results match with findings of studies using dynamic
testing in typical school populations (e.g., Resing et al., 2009; Resing, Tunteler, de Jong, & Bosma, 2009; Resing, 2000; Tzuriel, 2000a). Furthermore, the results extended studies concerning dynamic testing with children with intellectual disabilities, which focused either on severe intellectual disabled people (e.g. Büchel, Schlatter, & Scharnhorst, 1997; Hessels-Schlatter, 2002a) or adolescents (Lifshitz et al., 2005), as children in our study were 10-12 year old students with moderate to mild intellectual disabilities.

We also found that the use of the structured graduate prompts protocol clearly enabled us to measure the need for instruction. Results revealed that the number of prompts and type of prompts, given during the training, varied between children, as was also found in typically developing children (Resing, Tunteler et al., 2009), and in children with math difficulties (Bosma & Resing, 2010). Between groups of children with comparable standard pretest scores and between groups of children with comparable low or moderate IQ scores, we found that the number and type of prompts varied. This implies that based on standard IQ scores children are viewed as having comparable low cognitive abilities, whereas our findings indicate that for a considerable number of children with intellectual disabilities, reliance on standard IQ scores may lead to incorrect estimations of their educational needs.

In addition it was shown that trained children varied in their need for assistance on the construction task. Children with various levels of assistance were able to construct analogy problems and a small number of children were able to construct more than half of these analogy items independently. This task requests both an active attitude of the child and a more profound understanding of analogical reasoning, compared to the relative passive pre- and posttest task of solving analogy problems. The relation between the score on the construction task and the posttest score of the children was considerable, but not perfect. The addition of a construction task to the dynamic procedure may therefore be useful in providing supplemental information about a child’s potential for learning. In future research children could be asked to verbally explain their analogy, which may provide deeper insights in their understanding of analogical reasoning (Bosma & Resing, 2006). Also a comparison of the results on the
construction task between trained and untrained children is suggested to evaluate the effect of training on the construction task.

As we found differences in learning potential and variation in their needs for instruction and assistance, the opinions of the children’s special education teachers were of interest. Teachers considered the dynamic testing results and recommendations as valuable information, and in particular appreciated the information about type and number of prompts the children needed. The information was considered supplemental to the usual intelligence reports, and viewed as more positive, since the results focused on potential instead of deficits. These results are promising, since the potential of dynamic testing results are not only noticed by teachers in regular education (Bosma & Resing, 2008, 2010) but also by teachers in special education, who are expected to be quite familiar with and used to results of intelligence tests.

Although teachers’ verbal responses were positive regarding the results, repeated classroom observations did not show substantial changes in teacher-child interactions between the dynamically tested children and children of whom teachers only received standard reports. These findings in special education settings with small teacher-child ratios correspond to observational data gathered earlier in regular education (Bosma & Resing, 2008, 2010). Perhaps the time between providing information and observing in the classroom was too short for teachers to make changes (e.g. Witt, 1986). Inspection of the learning potential ratings by teachers showed a decrease from above average to average ratings for both trained and untrained children. The ratings related to IQ scores, but also to the total number of prompts children needed. Elaboration by teachers during the rating process might have been informative for the reason of the found changes. In general, the information in the reports might have been too new and different compared to what teachers were used too. Teachers may have had difficulty acting according to the recommendations and may have needed assistance, for example in classroom management, in translating the recommendations to their teaching practice.

Notable limitations of the study need to be mentioned. We realize that in this study we one-sided focused on the cognitive functioning of the children and did not take
into account the physical or behavioral difficulties these children experience, potentially those influencing their learning and classroom functioning. We did not find differences on the dynamic test results between children with or without a specific diagnosis, but of course comparing groups was hardly possible. Further, only a very small number of teachers were included in the study, which requires careful interpretation.

Our findings certainly have implications for the diagnostic assessment practices and placements, especially in special education. Placement for special education is highly based on outcomes of intelligence tests, but since IQ scores hardly inform about children’s learning processes (Haywood & Lidz, 2007) or instructional planning (Taylor, 1987; Ysseldyke, 1987), reliance on IQ measures has consequences for the teaching children receive. Children might achieve below their actual potential when assistance and instructions are not accommodated to their needs, consequently they are not able to profit from the standard instruction given. Further, underachievement may be established by the low expectations teachers gather of children’s potential based on the intelligence scores, which leads to limited challenging of children’s zone of proximal development (Feuerstein et al., 1979; Tzuriel, 2000a). The use of dynamic testing, with a graduated prompts technique, in addition to the known intelligence level, more accurately informs psychologists and teachers about the potential for learning for each child, their zone of proximal development, and, more specifically, the instruction children need to be successful learners.

We conclude that dynamic testing, including a graduated prompt training and a construction task provide additional quantitative and qualitative information about potential for learning of children with moderate to mild intellectual disabilities in special education. In particular, the variation in instructions children profit from can inform teachers and guide educational planning, which all might lead to unfolding the potential for learning of children, despite their severe learning difficulties. To reach this goal however, dynamic testing needs to become part of (school) psychologists’ education and in service training, and the concepts of dynamic testing and potential for learning should be incorporated in teacher education programs.