



Universiteit  
Leiden  
The Netherlands

## Dynamic testing in practice : shall I give you a hint?

Bosma, T.

### Citation

Bosma, T. (2011, June 22). *Dynamic testing in practice : shall I give you a hint?*. Retrieved from <https://hdl.handle.net/1887/17721>

Version: Not Applicable (or Unknown)

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

Downloaded from: <https://hdl.handle.net/1887/17721>

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

## **CHAPTER 6**

Teacher's appraisal of dynamic assessment outcomes:  
recommendations for weak mathematics-performers

The contents of this chapter are published in:

T. Bosma & Wilma C. M. Resing (2010).

*Journal of Cognitive Education and Psychology*, 9, 91-115.

## **Abstract**

This study investigated teachers' evaluations of reports and recommendations, based on outcomes of dynamic assessment, regarding their second grade pupils with math difficulties. Thirty one teachers and 116 pupils assigned to an experimental or control condition participated. Reports for children were based on administrated math and memory tasks and either a dynamic test (Seria-Think Instrument) or standard test (Raven PM). Teachers were observed, interviewed, rated the learning potential at two moments and evaluated specific dynamic assessment information in a follow up questionnaire. Results showed that teachers valued the dynamic assessment reports and recommendations overall as meaningful; as did teachers reading static reports. Learning potential ratings appeared to be affected by the reports. Dynamic assessment information and recommendations were valued as applicable for constructing individual educational plans; personal factors (seniority and teaching experience) appeared of influence. Dynamic assessment is recommended to be part of teacher's curriculum, to realize its potential.

## **Acknowledgements**

We thank A. Demir, M. Kooren and B. Leest for helping us collect and code data.

## Introduction

Primary aims of educational assessments are to evaluate current school achievement, predict future achievement and, possibly more importantly, prescribe educational interventions for children who do not show enough progress in school (e.g., Caffrey, Fuchs, & Fuchs (2008). Psychological assessment mostly ends with forms of classification, prediction of future functioning, identification of eligibility for therapy or special educational services, and clarification of, for example, learning difficulties. Educational psychologists, however, often lack information enabling them to handle prescriptively. They need an extra diagnostic phase to be able to transform assessment results into practical educational recommendations for teacher interventions for the particular child (Brown-Chidsey & Steege, 2005). One-point-in time, static assessment procedures that do not allow any form of feedback during testing, are not supposed to provide this information. Aim of the current study was to investigate how outcomes of dynamic testing could contribute to the prescriptive part of educational assessments.

During the last decades educational psychologists and other professionals in the field of psychological and educational assessment have noticed the increased demand by teachers to provide recommendations tailored to the needs of the individual child (e.g., Kubiszyn & Borich, 2009). Some of them are, for example, discussing Responsiveness to Intervention (RTI) as a method providing appropriate intervention for children with learning problems, including a more valid identification method for these groups of children (e.g., Fuchs et al., 2007). This better awareness to individual differences and educational needs seems to be related to the diversity in child problems in today's elementary school classrooms and to the request that teachers have to adapt their teaching to the various needs of particular children.

The necessity to adapt the curriculum and instruction by evidence-based methods to the needs of the students is expected to be part of the repertoire of special education teachers already, but is nowadays also expected in regular education.

## *Chapter 6*

Applying evidence based methods, such as RTI, assumes that at least most children would receive education they need in an early stage of their development, and not as a response to failing (Fuchs et al., 2007). In the Netherlands RTI is not yet practiced; however, educational psychologists have been trained to work according to a “needs bases assessment” model that emphasizes instructional recommendations (Pameijer, 2006). Both methods still lack the availability of good diagnostic instruments (Fuchs et al., 2007; Pameijer, 2006). It is assumed that outcomes of dynamic testing can fulfill a role in providing these recommendations for instruction, because dynamic testing provides exploration of the nature and amount of help, assistance, instruction a child needs to solve cognitive-intellectual problems and school achievement tasks (Grigorenko, 2009, Jeltova et al, 2007; Haywood & Lidz, 2007).

Purpose of the current study was to investigate teachers’ appraisal of outcomes of dynamic testing reports and recommendations regarding grade-2 children with math difficulties in a semi-diagnostic setting. In a former study we already investigated teachers’ opinions of and recommendations based on reports written on outcomes regarding dynamic testing. Results were promising but not always consistent, partly because the participating children were all typically developing and attended regular elementary schools. The study did not intend to imitate the diagnostic process as would be usually the case, and we had to conclude that teachers did not express their concerns to be addressed by the assessment procedure and did not find the recommendations very supplementary (Bosma & Resing, 2008).

In the current study, we concentrated on recommendations based on dynamic testing regarding elementary school children with severe math difficulties in their own classroom environment. We investigated (1) which type of information in the dynamic or static reports teachers valued as useful for their classroom practice and for formulating educational plans regarding individualized instruction to these children, (2) how estimated learning potential and classroom practice was affected by our dynamic reports, and (3) teacher’s opinions regarding the applicability of dynamic assessment information for guiding individual educational plans.

***Theoretical background***

In various studies the effects and applicability of dynamic testing in educational settings have been studied (e.g., Hessels-Schlatter, 2002a; Lidz, 2002, Resing 1997; Resing, Tunteler, De Jong & Bosma, 2009). Results revealed insights into children's potential for learning, their need for instruction and their responses to feedback. A promising dynamic assessment approach, in which feedback is provided according to a hierarchical structured prompts system, is the graduated prompts approach (e.g., Campione & Brown, 1987; Elliot, Grigorenko & Resing, 2009, Pena, 2000; Resing 1993, 2000). In this approach the provided feedback, both type and amount of feedback are supposed to be indicative for a child's potential for learning; in addition, qualitative elements would in principle be of use to guide classroom recommendations (Bosma & Resing, 2006; 2008; Haywood & Lidz, 2007). Results of a reversal task after dynamic testing revealed additional information to guide recommendations (e.g., Bosma & Resing, 2006). Nevertheless, dynamic tests have not been fully incorporated into educational psychologists' practice (Elliott, 2003; Grigorenko, 2009; Jeltova et al., 2007) for reasons of time constraint or lack of training opportunities (Deutsch & Reynolds, 2000; Haney & Evans, 1999), although advantages of dynamic testing have been acknowledged in the field of education. Two articles in which the value and application of dynamic assessment methods were investigated for educational psychologists (Deutsch & Reynolds, 2000) and special education coordinators (Freeman & Miller, 2000) revealed that respondents valued dynamic assessment, in particular the information about learning processes and strategies, the focus on strengths and the potential of dynamic assessment in linking assessment to educational interventions by providing hands-on information for teachers.

Psychological reports, including general outcomes of assessment and recommendations, would in principle be helpful means to gain insight into teacher preferences and expectations about the meaningfulness of dynamic assessment information. However, several studies showed that teachers appear to be critical readers, as far as psychological reports are concerned (e.g., Pelco, Ward, Coleman & Young, 2009) and, although report writing is an important professional role for

psychologists, the reports have not served teachers well up till now. Reports often lack the practical and concrete recommendations teachers expect (Haywood & Lidz, 2007), showing a gap between perceived usefulness by teachers versus psychologists (Hagborg & Aiello-Coultier, 1994) and between given recommendations and provided educational services (Kanne, Randolph & Farmer, 2008).

Studies comparing types of psychological reports have given more insight on readers and teachers preferences (e.g., Pelco, et al., 2009; Salvagno & Teglassi, 2001). Reports based on dynamic assessment potentially could include information to provide teachers with relevant ideas for individual planning, because these reports, among others, entail information regarding instructional strategies children could profit from (Delclos, Burns & Vye, 1993). A first study of usefulness of reports based on dynamic assessment was conducted by Hoy & Retish (1984). Contradictory to their expectations, standard reports were rated as more valuable by graduate students than dynamic reports based on Feuerstein's Learning Potential Assessment Device. Information for planning interventions was rated as insufficient on both report types. Unfamiliarity with the type of information in the report as well as its unusual format were reported as main reasons for these low rating of the dynamic reports. However, in a comparable study, Delclos et al. (1993) did find no report preference over teachers. Yet, prior teacher training in either direct or mediated learning did influence report preference. Teachers both trained in mediated learning and using a process oriented approach to planning interventions did notice implications for interventions more often, while teachers who did not had such a process oriented approach to prescriptive programming did notice implications but only if these were explicitly stated in the reports.

Planning interventions based on psychological reports appeared rather difficult for teachers. In a study of Pelco et al. (2009), about half of the participating teachers were not able to brainstorm over a single appropriate intervention after reading a psychological report. In our study we therefore stressed the importance of capturing teachers' preferences and needs regarding psychological reports, in particular regarding recommendations based on these reports. Results of Hulburt (1995) regarding the usefulness of assessment reports for planning interventions are of interest. She

investigated teachers' ratings of information regarding the applicability in planning interventions in preschool settings by comparing three types of reports. Information about teaching strategies and learning processes, in particular specified in the dynamic assessment reports, was rated as most valuable compared to all other information. Teachers further valued reported skills, information regarding monitoring progress, understanding a child's difficulties and practical recommendations regarding their pupil, which were all specified in both dynamic assessment and curriculum based reports. The diagnostic process, only described in the standard report, was also judged as important. Teaching experience or any type of teacher training did not influence the ratings. Results of an evaluation by special education coordinators in the UK regarding norm referenced, criterion referenced and dynamic assessment reports (Freeman & Miller, 2000) supported these findings. Although dynamic assessment reports were relatively unfamiliar to the coordinators, they rated the information as potentially useful in understanding student difficulties and in formulating individual educational plans.

### ***The present study***

Evaluating reports and recommendations regarding imaginary children has been mentioned as a limitation in studies regarding teacher preferences (e.g., Pelco et al., 2009). Therefore, like in our previous study we chose to use a practice based research design, in which we recruited teachers and their children with math difficulties in a simplified and (semi)controlled -diagnostic assessment process, observations and a written diagnostic report for each participating child. Including children with math difficulties in this study enabled us to formulate recommendations in response to actual present problems, as a referral question would have served in a diagnostic process.

To be able to provide recommendations regarding amount of needed type of instructions for every child and relate these to their arithmetic achievement, we made use of a dynamic test with graduated prompts techniques that has been developed as an adaptation of Tzuriel's (2000b) Seria-Think Instrument (Resing et al., 2009), in which problem solving, accuracy and seriation problems are the central focus. As children with math difficulties belong to a broad category including children with a variety of problems



in math, we extended our assessment with a general math achievement task, to provide descriptions of strengths and weaknesses in arithmetic, and two working memory tasks. Working memory capacity has been shown to influence math achievement results and efficiency (e.g., Swanson & Beebe-Frankenberger, 2004). We assumed that this assessment procedure would enable us to observe children's general task behavior and formulate recommendations based on the outcomes of the dynamic test in relation to the child's functioning on math and memory.

Our first aim was to examine teacher's ratings of the meaningfulness of reported information for planning individual educational plans. We elaborated on the previous study in which teachers - inconsistently - expressed their appreciation of the contents of the dynamic testing reports and regarded the recommendations as potentially applicable for their teaching practice. By focusing in the current study on children with moderate to severe math difficulties, we presumed that outcomes of assessment and recommendations would be of interest to teachers. Therefore and in line with the results found in studies related to the value of reported dynamic assessment (Delclos et al, 1993; Freeman & Miller, 2000; Hulburt 1995), we (1) expected to find differences in appreciations of the reports and recommendations between teachers in the experimental and control group. In particular, we expected teachers in the experimental group to value the reported information and recommendations higher as they received elaborated information about the learning process and instructions of their children in the dynamic reports.

Our second aim was to inspect how teacher estimations of learning potential would be influenced by the dynamic testing reports. We (2) expected that teacher ratings of learning potential, tapped after the reports were read, would relate stronger to the reported learning potential than comparable ratings before the experiment started. We did not expect an overall increase in learning potential ratings as children were expected to differ in their response to instruction, and actual learning potential. Delclos, Burns & Kulewicz (1987) reported that viewing a dynamic test report can affect teachers' expectations of children with learning problems, and Delclos et al. (1993) found that reading dynamic reports can lead to different expectations regarding a child's

learning potential compared to reading static reports of the same child. In the present study we further investigated whether we could observe changes in the classroom and in teacher-child interactions prior and post our assessment and reports. We decided, in contrast to our previous study, not to integrate the classroom observed information in the reports for teachers, because the essence of our reports then and in the current study was to compare and value information about learning potential in dynamic versus static reports.

The third focus of the study was to investigate teacher's opinions regarding the applicability of dynamic testing information for guiding individual educational plans. We also wanted to capture the opinion of teachers in the control group regarding reported information based on dynamic testing. We (3) expected teachers in the control group to rate the report sections as more valuable. In addition we explored the overall ratings of teachers of specific and isolated excerpts of dynamic testing results and we expected teachers to rate these results as informative and useful in setting up individual plans. We based this hypothesis on Freeman and Miller (2001) who found that special needs coordinators rated information regarding type and regarding amount of instructions a child needs, improvement after interventions and specific strategies, as useful for planning interventions.

A fourth aim of this follow up was to obtain an overall evaluation of the reports and recommendations. We (4) expected teachers in the experimental group to value the individual reports and recommendations higher than teachers in the control group.

## **Method**

### ***Participants***

Participants in this study were 116 second grade elementary school children (78 girls and 38 boys)<sup>1</sup> with a mean age of 8;0 years ( $SD = 5.9$  months; range 86 - 115 months). All children were recruited from 23 schools in large cities of the Western part of the

---

<sup>1</sup> The unequal numbers were the actual results of our recruitment of children as we'll explain in our design. Apparently more girls than boys achieved low grades in mathematics, at least in this grade.

Chapter 6

Netherlands. About half of the children had an indigenous Dutch background, while the other half of the children had an ethnic minority background, with parents born in the Dutch Antillean, Turkey, or Morocco. The in 31 participating teachers (26 females; 5 males) had a mean age of 38;1 years ( $SD= 10;6$  years; range 23 - 56 years) and their teaching experience ranged from 2 to 25 years ( $M= 12;9$  years and  $SD= 7;6$  years). One teacher was ill at the end of the experiment and she and her 4 pupils have been taken out of the analyses. On the follow up questionnaire 21 teachers (68%) responded. Written parental consent was obtained for all participating children.

**Design**

Schools and teachers were selected first in this study. Participation of their children was based on low math achievement scores (last 20% on Cito monitoring test for mathematics, Janssen, Scheltens & Kraemer, 2005) and no prior grade retention or repetition. Both children and their teachers were matched to an experimental or control condition, based on the children’s gender, age and Raven PM score (Raven, Court, & Raven, 1979), and teacher’s regularity of reading psychological reports. The Raven served as a global measure of the children’s general level of cognitive-intellectual abilities. The administrative order of various measurements including the distribution of teachers and children over the two conditions has been displayed in Table 6.1.

**Table 6.1.** Design of study

Condition	Teacher measures I	Pupil assessment		Report	Teacher measures II	Follow up
		pretest	Training	Posttest		
<b>Experimental</b> 14 teachers 57 pupils	Interview observations Checklist	Raven	Graduated prompt training	Dynamic	Interview Observations checklist	Standard sample report questionnaire
		Dynamic test		Dynamic test		
<b>Control</b> 17 teachers 59 pupils		Math & Working-Memory	-	Math & Working-Memory		Dynamic sample report questionnaire

Teachers in both conditions were interviewed regarding their educational practice and construction of individual educational plans, filled out a checklist about school behavior and learning potential of the children involved in the experiment, and were observed on teacher-child interactions, use of feedback and task instructions during math lesson. Math and memory tasks were administered to obtain report measures for all children and this was repeated several weeks later. Then, only the children in the experimental group were tested dynamically, consisting of pretest, training and posttest sessions. Children in the control group received the pre- and posttest, without training. Next, teachers in the experimental group received reports regarding outcomes of the dynamic test, and math and memory task results, whereas in contrast, teachers in the control group were given standard reports, in which the results of the Raven, math, and memory tasks were described. To evaluate the value of the reports and recommendations, all teachers were interviewed a second time, parallel versions of the school behavior checklist were filled in, and classroom observations were conducted to observe differences in practice as a consequence of provided recommendations. To evaluate the contribution of reports based on outcomes of dynamic testing all teachers were given a follow-up questionnaire with a different report attached, either a standard report (experimental condition), or a dynamic report (control condition).

### ***Instruments***

*Raven's Progressive Matrices (Raven PM).* To obtain an indication of each pupil's general level of cognitive-intellectual ability before the experiment started, the Standard black and white version of the Raven PM (Raven, Court, & Raven, 1979) was administered. Raw scores were used for blocking participating children over the conditions and individual results of level of cognitive functioning were interpreted and reported in the standard control-group reports.

*Memory tasks.* To obtain a second measure for the psychological reports, scores of all children on two auditory working memory tests were collected: the *Digit Span backwards* subtest of the WISC-III<sup>NL</sup> (Wechsler, 2005), and the *Auditory Digit Sequencing* subtest of the Swanson-Cognitive Processing Test (S-CPT, Swanson, 1995). Correct recalled items

## Chapter 6

were scored and categorized; for Digit Span a recall of three numbers was interpreted as average for this age; for Auditory sequencing two to three correct recalled items was considered as average for this age group (Swanson & Beebe-Frankenberger, 2004).

*Arithmetic tests.* The *CITO monitoring test for mathematics* (Janssen, Scheltens & Kraemer, 2005), measuring knowledge and skills regarding to numbers, calculations, ratio, fractions, measuring, time and money, was used to determine children with math difficulties, in particular children who's scores fell in the lower 20 percent. Results of the test were gathered by the schools during the middle of the school year. From the *Dutch Didactic Age test for arithmetic/math* (De Vos, 2001) the first 4 pages of arithmetic/math problems for children grade 1 and 2 were administered. Problems increased in difficulty, ranging from simple counting the pictured beads, read clocks, solving money and measurement problems and simple calculations. Children were instructed to answer as many of the items as they could. Time was unlimited, and ranged from 10 to 30 minutes. Scores could range from 0-105; the total was used to indicate the math grade level.

*Dutch School Behavior Checklist Revised (SCHOBL-R).* In order to measure typical classroom behaviors of children, the SCHOBL-R (Bleichrodt, Resing & Zaal, 1993; Resing, Bleichrodt & Dekker, 1999) was completed by all teachers, including the factor learning potential (Bosma & Resing, 2008) to capture teachers' estimation of learning potential.

*Seria-Think Instrument.* The Seria-Think Instrument is a dynamic test to measure both seriation and early math skills (Tzuriel, 1998, 2000b). The test consists of a wooden cube, with holes in various depths, a series of red rods and a measuring rod, and the child has to insert a series of rods so that these rods are of increasing, equal or decreasing height. A good strategy to solve the seriation problems is a measurement strategy. Insertion behavior, therefore, has to be minimized while the use of the measuring rod is unrestricted. In the manual the pre and posttest instructions are well defined while the instruction for the training is not specified into details, but have to be given according to the principles of mediation (Tzuriel, 1998). Following Resing, De Jong, Bosma & Tunteler (2009) and Resing, Tunteler et al., (2009) we developed for the five by five version of the Seria-Think Instrument a training based on the graduated prompts approach (e.g.,

Campione & Brown, 1987). This training consisted of standardized protocols with prompts for every action a child undertook to complete a series of rods. These prompts were hierarchically structured, starting with general meta-cognitive prompts (e.g., “what do you have to do?”), to more concrete, cognitive task-specific prompts (e.g., “how long does your rod needs to become?”) to modeling prompts in which the experimenter showed the pupil how to measure depth, height or select a rod (e.g., Resing, 1993, 1997). Measures for learning potential were operationalized in terms of the levels of insertions and measurements after training, and combined with the minimum number of prompts children needed to complete series of rods on his or her own, until a fixed learning criterion was reached (e.g., last row without help). Number and type of prompts were viewed as providing central information in formulating recommendations for the experimental, dynamically assessed group of children. Reported learning potential of each child was given in one of five categories, based on the amount of needed hints (low-middle-high), the number of measurements (low-middle high) and number of insertions (low middle high).

*Reports and Recommendations.* As in an assessment, reports including recommendations regarding encountered math difficulties and need for instructions were written for all children. In addition, a general description of the assessment was provided to all teachers, while teachers in the experimental group additionally got an explanation of the dynamic test. Although the reports differed in content for children in both conditions, the reports were build up according to the same structure and order. First the outcomes of the assessment were mentioned, including observations during testing, math and memory task results, including encountered difficulties and the learning potential estimations of the teachers.

<p><i>Results:</i></p> <ul style="list-style-type: none"> <li>• Observations (task attitude, openness, motivation, concentration)</li> <li>• Seria-Think pretest and posttest: Number of Insertions and measurements</li> <li>• Graduated prompt training: amount and type of hints (meta-cognitive/ specific)</li> <li>• Outcomes of math and memory tasks I and II, and encountered difficulties</li> <li>• Results of the teacher checklist I and teacher’s estimated learning potential (interview I)</li> </ul> <p><i>Recommendations:</i></p> <ul style="list-style-type: none"> <li>• Amount of instruction needed</li> <li>• Type of needed instruction</li> <li>• Examples of instruction related to the encountered arithmetic problems</li> <li>• Needed practice/type of arithmetic problems (e.g. automatization)</li> <li>• Adaptations (tempo &amp; task behavior)</li> <li>• Need for positive feedback, support, challenge</li> </ul>	<p><i>Results:</i></p> <ul style="list-style-type: none"> <li>• Observations (task attitude, openness, motivation, concentration)</li> <li>• Raven</li> </ul> <ul style="list-style-type: none"> <li>• Outcomes of math and memory tasks I and II, and encountered difficulties</li> <li>• Results of the teacher checklist I and teacher’s estimated learning potential (interview I)</li> </ul> <p><i>Recommendations:</i></p> <ul style="list-style-type: none"> <li>• Level of reasoning and needed math practice (a lot / a little)</li> </ul> <ul style="list-style-type: none"> <li>• Needed practice/type of arithmetic problems (e.g. automatization)</li> <li>• Adaptations (tempo &amp; task behavior)</li> <li>• Need for positive feedback, support, challenge</li> </ul>
--	--

**Figure 6.1.** Construction of the individual dynamic and standard psycho-diagnostic reports

Than, the child’s functioning was explained. The report finished with recommendations regarding needed practice of specific encountered math and, if present, memory difficulties, adaptations in the curriculum (frequent repetitions, visual supports, extended time), and needed approach (positive feedback, boost confidence) were given.

The main difference between reports regarding experimental versus control-group children was that, while reports for children in the control-condition described levels of general cognitive functioning based on the Raven, reports of children in the experimental condition prominently outlined the need for and response to hints, as well

as the type of hints, achievement of the learning criterion and application of learned problem-solving skills. See Figure 6.1 for an overview of the results and recommendation section. Total reports were one to one and half page long.

*Teachers' Interview.* Teachers were interviewed before the start of the experiment. The structured interview questions were focused on teachers' practical experience with psycho-diagnostic reports, writing individual educational plans, and types of recommendations preferred in reports. For each participating pupil his or her teacher before testing took place rated the child's learning potential on a 6-point scale. In a second structured interview, that took place four weeks after provision of reports, the 4 main elements of information given in all reports were stressed: a) task behavior and observations of the child, b) a child's need for instruction, c) the child's cognitive functioning, and d) results on the math and memory tasks. In addition, teachers rated on a 4-point scale (very little-little-many-very many) the usefulness of these types of reported information in regards to constructing individual educational plans, providing insights into cognitive functioning and getting insights into the instruction the child needed. Further, teachers rated both the general usefulness and applicability, and again the child's learning potential on a 6-point scale.

*Observations.* Parts of Lidz' Observing Teaching Interactions, as far as they did relate to dynamic testing, were used to observe the teacher's teaching practices: Intentionality (involvement of teacher), Task Regulation (type of task instruction), Praise and Feedback (frequency of positive feedback), Challenge (challenging children's ZPD), Informing Change (informing the child's achievement) and Contingent Responsivity (response to and balancing of children's needs) (Lidz, 2003). Teacher-child interaction in the classroom was observed as well. Parts of the Mediated Learning Experience Rating Scale, developed by Lidz (1991, 2003; in translation Van der Aalsvoort (1994) were used. Except for Contingent Responsivity, the components on this scale were similar to those on the Observing Teaching Interaction Scale. All observations took place in the classroom during math instruction and required at least half an hour. To acquire a good sense of the



## Chapter 6

classroom atmosphere and to practice the use of observation scales, experimenters were supervised during their first couple of observations.

*Follow up counter reports and questionnaire.* Six weeks after the second series of interviews, observation and checklists were finished, there was a follow up session. A sample dynamic testing report and a sample standard report of a fictional child were constructed resembling the assessment reports we had written and handed to teachers concerning each participating child. These so called 'counter-reports' were used to capture control-condition teacher's opinions about dynamic testing reports and information. They were asked to study the report and answering the questionnaire afterwards. All teachers rated the possible contribution of (1) the results and recommendations, and (2) the usefulness of dynamic testing information (type of instruction, amount of instruction, improvement after hints, and problem solving strategies), based on items Freeman & Miller (2001) used in their questionnaire. Finally teachers were asked to rate the usefulness of the reports and recommendations they had received for their children and if they would be able, based on the individual reports and recommendations, to make changes in their teaching practice in future.

### **Procedure**

Before the experiment started the Raven PM was administered in small groups and teachers were asked to fill in the SCHOBL-R checklist for all participating children. In addition, experimenters (three psychology master students) conducted a 10-minute during structured interview with the teacher and did classroom observations. Then, all children were assessed with the Dutch Didactic Age test arithmetic/math (De Vos, 2001) and two, individually administrated, working memory tasks: subtest Digit span backwards (Wechsler, 2005) and the auditory digit sequencing subtest (Swanson, 1995). Next, the Seria-Think instrument was administered individually in three weekly 20-30 minute sessions to the experimental group; the children in the control group only received the static pre- and posttest. A week later the math and memory tasks were repeated to all children. Individual written reports, including recommendations, were

handed to each teacher regarding each participating child. Reports with specific recommendations based on the dynamic test were provided only for children in the experimental condition. All reports were under our supervision written by the experimenters and edited to insure similar content, structure, and style. Four weeks later teacher checklists were handed out again, and classroom observations took place again. In a second interview teachers were asked to evaluate the recommendations and to provide again learning potential ratings. In a follow up the last month prior to the summer holidays a 'sample'-report and evaluation questionnaire were handed to the teachers.

## Results

Before reporting our findings regarding the research questions, we examined whether experimental and control groups did significantly differ on their mean level of cognitive functioning, age, math, gender, and regarding the mean age of the teachers and teachers' mean experience with report reading. One-way ANOVA's with condition as independent variable and child's age, Raven PM score, and the score on CITO-Math as dependent variables, did not show significant differences between the two conditions on any of these variables. Additional  $\chi^2$  analyses with condition as independent variable and child's gender and background as dependent variables demonstrated no significant differences between conditions either. The majority of teachers had some experience with reading psychological reports: 61% read a report more than three times a year, 26% once or twice a year, and 13% did not see a report at all. Analyses ( $\chi^2$ ) revealed no significant differences in report reading experience between teachers in both conditions.

Because we studied a specified group of children with math difficulties in elementary mainstream education, we first examined the assessment results regarding the math and memory tasks and the dynamic test as well. Scores on the Raven PM ranged from 8 to 43 with no significant differences between  $M_{\text{exp}} = 24.84$  ( $SD = 8.55$ ) and  $M_{\text{control}} = 25.86$  ( $SD = 8.10$ ). Scores on the didactic arithmetic/math test ranged from 11-

62 with no significant differences between  $M_{\text{exp}} = 32.70$  ( $SD = 10.42$ ) and  $M_{\text{control}} = 31.58$  ( $SD = 9.61$ ) respectively. On the memory tasks children achieved comparable scores with  $M_{\text{exp}} = 3.54$  ( $SD = 1.14$ ) and  $M_{\text{control}} = 3.59$  ( $SD = 1.25$ ) on the Digit Span Backwards and  $M_{\text{exp}} = 1.75$  ( $SD = .93$ ) and  $M_{\text{control}} = 1.78$  ( $SD = 1.29$ ) on Swanson Auditory Sequencing Task respectively. It can be concluded that both teacher and pupil conditions were equally matched on the variables above.

Further, to be able to make descriptive classifications on number of hints and other dynamic test outcomes, we analyzed the results of the dynamic test. We inspected both change in number of insertions (decrease) and number of measurements (increase). Table 6.2 gives an overview of the means and standard deviations for both experimental and control group children on these two measures.

**Table 6.2** : Means and standard deviations of number of insertions and number of measurements at pretest and posttest for the experimental and control group.

	<i>Experimental group</i>		<i>Control group</i>	
	<i>pre</i>	<i>post</i>	<i>pre</i>	<i>post</i>
<i>Number insertions</i>				
Mean	87.44	60.6	101.57	83.24
SD	32.19	26.73	47.50	36.22
<i>Number of measurements</i>				
Mean	7.30	20.30	6.71	8.36
SD	7.56	9.76	7.23	8.42

Analyzing the dynamic test data, by performing a multivariate repeated measures (RM) analysis of variance with session (pretest – posttest) as within subjects and condition (experimental – control group) as between subjects factors and number of insertions and measurements as dependent variables revealed significant effects for session (*Wilks's*  $\Lambda = .55$ ,  $F(2,112) = 46.84$ ,  $p < .001$ , partial  $\eta^2 = .46$ ) and session x condition (*Wilks's*  $\Lambda = .73$ ,  $F(2,112) = 20.65$ ,  $p < .001$ , partial  $\eta^2 = .27$ ). Univariate analysis revealed

a significant session x condition effect for number of measurements ( $F(1,113) = 41.04$ ,  $p < .001$ , partial  $\eta^2 = .27$ ), but no significant interaction effect for insertions ( $p = .312$ ). Pupils in the experimental group increased their number of measurements significantly more from pretest ( $M = 7.30$ ,  $SD = 7.56$ ) to posttest ( $M = 20.30$ ,  $SD = 9.76$ ) than children in the control group ( $M = 6.70$ ,  $SD = 7.2$  and  $M = 8.36$ ,  $SD = 8.41$ ). During the training session with 5 series with 5 items of problems to solve, children needed a considerable number of prompts ( $M = 28.80$ ,  $SD = 19.66$ ), which varied between 1-97 prompts needed.

To answer the hypothesis that teachers in the experimental group would value the reports and recommendations higher than teachers in the control group, we investigated differences in appraisal, including recommendations, of teachers in the experimental versus control condition. Overall results showed that the majority of teachers valued the reported types of information as positive and helpful for all three aspects. In Table 6.3 the relative frequencies of the teacher evaluations over the different categories are shown on the left and the mean and standard deviations on the right. The table shows that teacher evaluations in both the experimental and control group were almost similarly distributed over categories a and c, but not over category b. Inspection of the data by  $\chi^2$  analysis showed a slightly different distribution (trend:  $\chi^2(3) = 5.76$ ,  $p = .08$ ) on reported observations and task behavior. Teachers in the control condition tended to rate this information as either little or much useful, whereas teacher ratings in the experimental group seemed to have more varied evaluations. In addition, a Mann-Whitney U tests was conducted for each type of reported information within the three categories. Data were analyzed on a one-tailed (.10) significance level, because it was our expectation that teachers in the experimental group would rate the value of the information higher than teachers in the control group but not vice versa. Regarding the category understanding cognitive functioning the Mann-Whitney U tests revealed a significant difference on both reported observation and task behavior ( $z = -1.44$ ,  $p = .074$ ), and reported instruction ( $z = -1.44$ ,  $p = .088$ ). Regarding insights in the needed instruction a significant difference was found on observation and task behavior ( $z = -1.52$ ,  $p = .08$ ). Teachers in the experimental condition tended to rate the reported

observation and instruction as slightly more positive than teachers in the control group regarding understanding cognitive functioning and needed instruction, although the reported differences must be seen as trends. For constructing individual educational plans, no differences were found between the two conditions.

**Table 6.3.** Distribution of teacher' ratings of meaningfulness of reported information (in percentages) and mean ratings and standard deviations regarding constructing individual plans, understanding the cognitive functioning and acquiring insights in the type of needed instruction

<i>Type of information</i>	<i>Condition</i>	<i>Distribution of relative ratings of Meaningfulness</i>				<i>Ratings of Meaningfulness</i>	
		<i>Very little</i>	<i>little</i>	<i>much</i>	<i>Very much</i>	<i>Mean</i>	<i>SD</i>
<i>For constructing individual plans</i>							
Observations and Task behavior	Experimental	6.5	12.9	19.4	6.5	2.57	.94
	Control	9.7	12.9	22.6	9.7	2.59	1.00
Type of instruction	Experimental	3.2	9.7	22.6	9.7	2.86	.86
	Control	6.5	12.9	35.5	0	2.53	.72
Cognitive functioning	Experimental	6.5	12.9	19.4	6.5	2.57	.94
	Control	9.7	12.9	22.6	9.7	2.59	1.00
Math and memory results	Experimental	6.5	3.2	25.8	9.7	2.96	.95
	Control	3.2	6.5	38.7	6.5	2.88	.70
<i>For understanding cognitive functioning</i>							
Observations and Task behavior	Experimental	3.2	9.7	25.8	6.5	2.79	.80
	Control	0	29.0	25.8	0	2.47	.51
Type of instruction	Experimental	3.2	6.5	29.0	6.5	2.86	.77
	Control	3.2	19.4	32.3	0	2.53	.62
Cognitive functioning	Experimental	3.2	12.9	22.6	6.5	2.79	.83
	Control	3.2	12.9	35.5	3.2	2.71	.69
Math and memory results	Experimental	3.2					
	Control	6.5	6.5	29.0	6.5	2.86	.77
	Experimental		12.9	29.0	6.5	2.65	.86
	Control						
<i>For insights in needed instruction</i>							
Observations and Task behavior	Experimental	3.2	9.7	29.0	3.2	2.71	.73
	Control	9.7	19.4	25.8	0	2.29	.77
Type of instruction	Experimental	6.5	9.7	19.4	9.7	2.71	.99
	Control	9.7	16.1	25.8	3.2	2.41	.87
Cognitive functioning	Experimental	3.2	12.9	25.8	3.2	2.64	.75
	Control	9.7	9.7	32.3	3.2	2.53	.87
Math and memory results	Experimental	0	9.7	25.8	9.7	3.00	.68
	Control	6.5	12.9	32.3	3.2	2.59	.80

Two-third of the teachers in the experimental group rated the reports as meaningful to very meaningful, against only half of the teachers in the control group. In the experimental group, one teacher (male, 28 years old) explained that he “received a lot of new information”; another teacher (female, 34 years old) stated: “Now I have much more information about what this child needs, instead of the failure analysis we get out of our curriculum tests”. Teachers in the control group with a positive rating explained overall that the reports and recommendations were a confirmation of what they already knew about their children’s functioning. Negative ratings (“little meaningful”) were given by 10-15% of the teachers in both conditions. Teachers in the experimental group explained that the reports and recommendations did not include enough new information regarding particular children. Neutral responses were given by 21% of the experimental-group and 40% of the control-group teachers.

Teachers’ responses were more conservative regarding possibilities to make changes in their teaching based on the reports. Not at all too little applicability was chosen by one-third of the teachers in the experimental group and fifty percent in the control group. However, most of these teachers argued that they had not had enough time to actually apply the recommendations in their practice. As one teacher in the experimental group (female, 34 years old) explained: “not yet, but next year we certainly will integrate these recommendations”. Most teachers responded neutral (57% experimental, 41% control). Teachers explained that they, for example, partially had implemented the recommendations or that they had become more conscious of their teaching method and instructions for a particular child. A teacher of the experimental group (female, 23 years old) said: “I have become more conscious about children who are weak and have adjusted my approach by, for example, repeating instructions”. Positive responses came from 12-14 % of the teachers in both conditions. One teacher in the experimental group (female, 50 years old) explained: “for J. I adjusted the curriculum and instructions based on the recommendations”, while another teacher explained that he did not yet had the time, but that he would implement the results. Our results, in terms of percentage of teachers evaluating reported information and

overall reports and recommendations, did not reveal clear differences between evaluations of teachers in both conditions, although the qualitative data gave a somewhat different perspective.

To investigate whether teacher ratings regarding learning potential as measured as part of a school behavior checklist differently changed as result of our experimental intervention, a RM analysis of variance was performed with total sum of items on the learning potential scale as within-subjects variable measured in two (experimental versus control) conditions specified as between-subjects factors across sessions (pretest and posttest). Our focus was on the interaction between condition and session. The analysis did not reveal such a significant interaction effect. Ratings in both conditions were low to below average and decreased slightly from first rating ( $M_{\text{exp}} = 33.28, SD = 11.47$ ;  $M_{\text{control}} = 31.65, SD = 12.45$ ) to the second rating four weeks after the reports were read ( $M_{\text{exp}} = 29.11, SD = 11.14$ ;  $M_{\text{control}} = 29.12, SD = 10.62$ ). Inspection of the relation between LP-checklist and LP-interview revealed moderate correlations (Pearson) in both conditions ( $r_{\text{exp}} = .57$ ;  $r_{\text{con}} = .41$ ) before the intervention took place, but differed at the second measurement between the two conditions ( $r_{\text{exp}} = .27$ ;  $r_{\text{con}} = .72$ ), being stronger for the ratings by teachers in the control condition, and vice versa. While we reported that the ratings of the LP-checklist slightly decreased in both conditions, the LP-interview ratings, in contrast, increased over time for both conditions ( $M_{\text{exp}} = 3.12, SD = 1.16$ ;  $M_{\text{control}} = 2.89, SD = 1.03$ ; and  $M_{\text{exp}} = 3.23, SD = 1.1$   $M_{\text{control}} = 3.25, SD = 1.09$ ) for first and second interview rating respectively. This was shown as a significant main effect resulting from another RM analysis of variance with the LP-interview rating as within-subjects variable measured in two (experimental versus control) conditions specified as between-subjects factors across sessions (pretest – posttest):  $F(1,111) = 7.90, p = .006, \text{partial } \eta^2 = .07$ . Teachers in both experimental and control group might have changed their direct estimation of the learning potential of their children when focused on it in an interview.

We further inspected the results of crosstabs from learning potential ratings prior to and post assessment and reports. The total of rated learning potential by the teachers was categorized into five categories, based on previous data on this factor.

**Table 6.4.** Comparison of reported and rated learning potential at pretest

		reported LP				
		Below Average	Low average	Average	High average	Above average
Schobl A rating	low	3 6.4%	1 2.1%	2 4.3%	0 .0%	0 .0%
	Low average	8 17.0%	3 6.4%	5 10.6%	5 10.6%	4 8.5%
	Average	3 6.4%	1 2.1%	4 8.5%	2 4.3%	3 6.4%
	High average	0 .0%	1 2.1%	1 2.1%	0 0%	1 1.8%
	high	0 0%	0 .0%	0 .0%	0 .0%	0 .0%
	Total %	29.8	12.8	25.5	17.0	14.9

**Table 6.5.** Comparison of reported and rated learning potential at Posttest

		reported LP				
		Below Average	Low average	Average	High average	Above average
Schobl B rating	low	1 2.1%	1 2.1%	0 .0%	0 .0%	0 .0%
	Low average	7 14.9%	1 2.1%	4 8.5%	4 8.5%	1 2.1%
	Average	6 12.8%	3 6.4%	6 12.8%	2 4.3%	4 8.5%
	High average	0 .0%	1 2.1%	2 4.3%	0 .0%	2 4.3%
	high	0 .0%	0 .0%	0 .0%	2 4.3%	0 .0%
	Total %	29.8	12.8	25.5	17.0	14.9

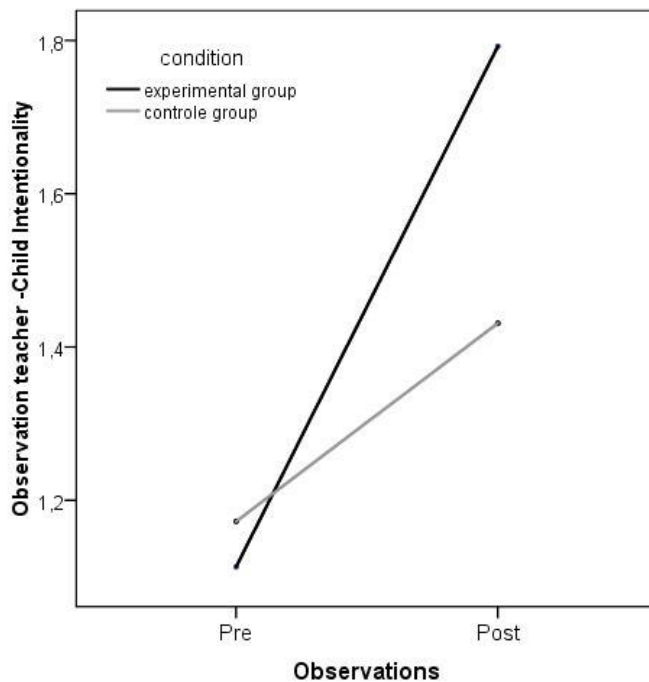
Similar categories were made for the description of the level of learning potential as it was written in the report (below average, low average, average, high average and high). We expected that teacher ratings of learning potential would be affected by reading the reports and recommendations and would change towards the reported learning



potential. Comparison of the crosstabs (see Table 6.4 and 6.5) of rated and reported learning potential shows that prior to the reports the teachers rated two-third of children as having below or low average learning potential, and only 5% of the children was rated as having above average learning potential.

The actual, reported learning potential was distributed more equally: less than half of the children fell in the categories below to low average and one-third in the high average to above average learning potential categories. After the assessment and reports, teacher ratings changed in the direction of the reported learning potential. Only 40% of the children were rated at a below or low average learning potential, whereas now 15% of the children received a rating at high average or above average learning potential.

We further examined whether changes in teacher-child interactions in the classroom, before and after the intervention, could be observed. RM analyses of variance with 6 subscales of Lidz' teaching-interaction rating scale specified as within-subject variables, measured in two (experimental versus control) conditions specified as between-subjects factors across sessions (pretest – posttest) revealed no interaction effects on any of the observed variables. Significant main effects were found for four of the subscales: Intentionality ( $F(1,29) = 5.90, p = .022, \text{partial } \eta^2 = .17$ ), Task behavior ( $F(1,29) = 19.40, p < .001, \text{partial } \eta^2 = .40$ ), Praise ( $F(1,29) = 7.00, p = .013, \text{partial } \eta^2 = .19$ ), and Contingent Responsivity ( $F(1,29) = 4.20, p = .05, \text{partial } \eta^2 = .13$ ). These results indicate that teachers in both conditions were observed to perform more intentional behaviors, task regulating activities, providing more positive feedback and showing more balancing of children's different needs. Besides examining teacher's general teaching practices by using these rating scales, we also investigated whether teacher-child interactions, if observed during math instruction, were differently affected by our reports.



**Figure 6. 2:** Pre and post intervention scores for observed teacher child interaction on Intentionality

RM analyses of variance with the Lidz' mediated learning experience rating subscales specified as within-subject variables, measured in two (experimental versus control) conditions specified as between-subjects factors across sessions (pretest – posttest) revealed a significant interaction effect for Intentionality  $F(1,109) = 5.01, p = .027$ , partial  $\eta^2 = .04$ ), and as can be seen in Figure 6.2, the observed intentionality increased more for the experimental group. Significant interactions effects were also found for Praise ( $F(1,109) = 9.07, p = .003$ , partial  $\eta^2 = .08$ ), and Challenge ( $F(1,109) = 12.88, p = .001$ , partial  $\eta^2 = .11$ ), but not in the expected direction as can be seen in Figure 6.3, A similar interaction pattern was found for Challenge as can be seen in Figures 6.4. On the subscale Task behavior the analysis only revealed a significant main effect,  $F(1,109) = 24.67, p < .001$ , partial  $\eta^2 = .19$ . The number of task regulating activities increased in both groups.

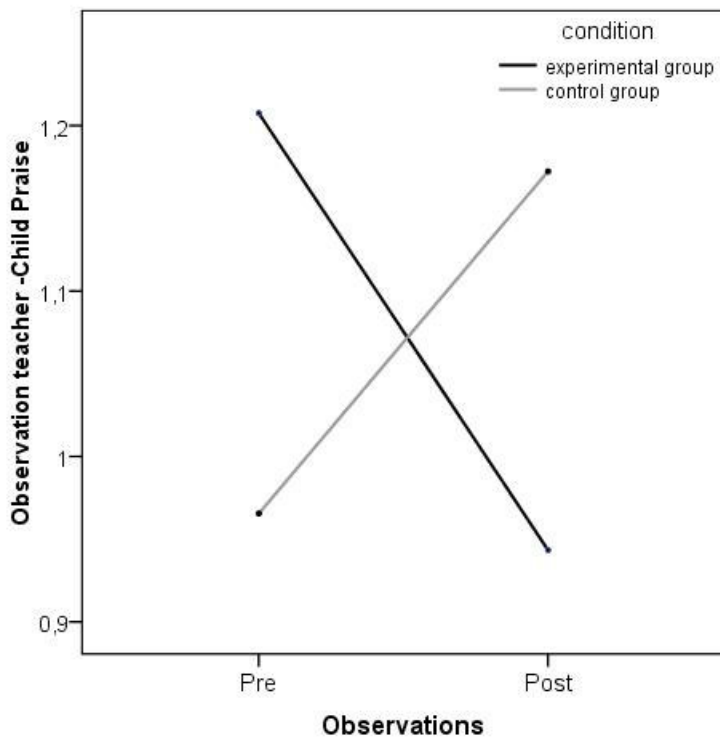


Figure 6.3. Pre and post-intervention scores for observed teacher child interaction on Praise

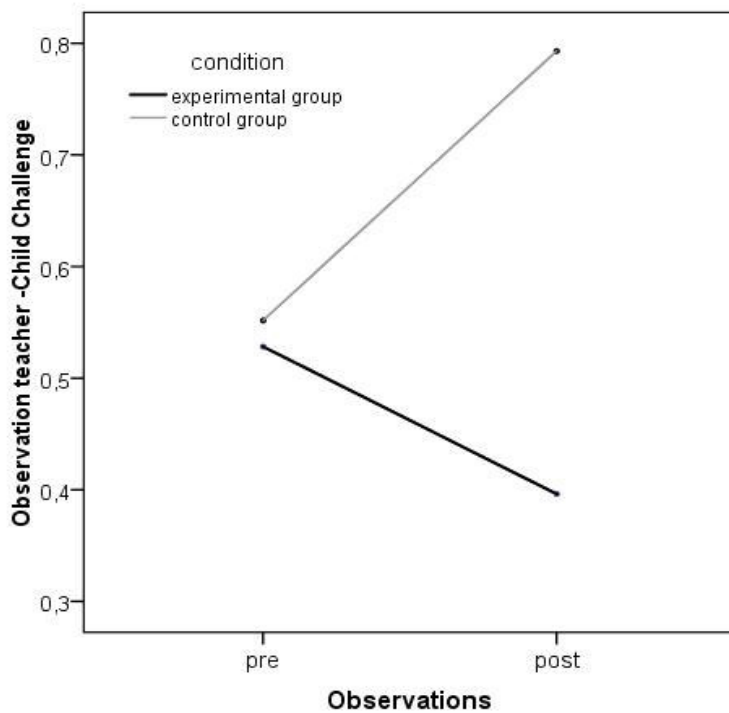


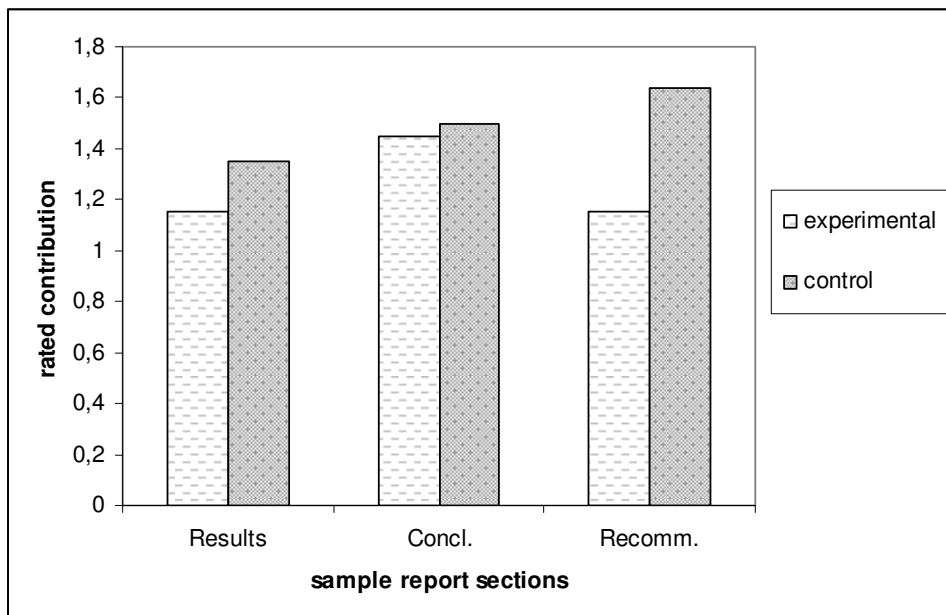
Figure 6.4. Pre and post intervention scores for observed teacher child interaction on Challenge

The follow up questionnaire regarding the practical value of dynamic testing for the teaching practice was filled in by two-third of all teachers. To inspect the surplus value of the sample reports compared to the assessment reports teachers had received earlier, an ANOVA was conducted for the ratings of the results, conclusion, and recommendations sections of the report.. Results showed that the control group who read this time a dynamic sample report, assigned relative higher values to the reports compared to the experimental group which received a standard report. Only a trend was found for the recommendation section  $F(1,19) = 3.99, p = .06$ . In Figure 6.5 the rated values are depicted for both groups on all three report sections.

The *type of instruction* in the reports was valued on average as 1.62 ( $SD = 1.02$ ) when teachers were asked about the usefulness for formulating individual educational plans. Eighty percent of the teachers valued this type of information as reasonable to good to excellent, *Amount of instruction* need by the child was rated on average as 1.39 ( $SD = .850$ ) and valued as reasonable or good to excellent by a large majority of the teachers, while only 16% of them regarded this information as little contributing to the construction of educational plans. The mean rating of the child's *improvement after hints* was 1.11 ( $SD = .66$ ) and was valued as 'reasonable or good' by 84% of the teachers and as little by 16% of the teachers. Finally, *practiced problem solving strategies* of the child was rated with a mean of 1.47 ( $SD = .77$ ). It was valued as 'good' to 'excellent' by two-third of the teachers, whereas one-fifth considered the information as reasonable contribution and 16% of the teachers did value it as 'little' contributing for constructing individual plans. Overall, it can be concluded that the majority of teachers did value the usefulness of information that comes forth out a dynamic testing, especially for constructing individual plans.

Teachers in both conditions valued the recommendations as average to above average with  $M_{\text{exp}} = 4.11$  ( $SD = 1.23$ ) and  $M_{\text{control}} = 3.73$  ( $SD = 1.27$ ). Extreme low ratings (1) or high ratings (6) were not given by any of the teachers and the results of an ANOVA with condition as factor and rated meaningfulness as dependent variable did not show significant differences between the two conditions. Overall, the teachers did

experience the reports and recommendations as useful. Six weeks later, teachers responded mostly neutral and said that application of recommendations by them would be possible, but that they had not yet the time to do so. RM analyses of variance with rating scale specified as within-subject variables, measured in two (experimental versus control) conditions specified as between-subjects factors across sessions (interview II – follow up) revealed no interaction effect. Significant main effects, however, revealed that teachers in both conditions rated the application of the recommendations more positive  $F(1,18) = 4.90, p = .04$ , partial  $\eta^2 = .21$ .



**Figure 6.5.** Rated value of the sample report sections per condition during Follow up

Because we did not find difference between conditions we explored whether age and experience of teachers played a role in the evaluations of dynamic assessment information, in the reports in general, and in the possibility to apply the recommendations in their daily teaching. As expected, age and teaching experience were highly related ( $r = .84$ ). We categorized teachers according to their age (young/old)

with a cutoff score of 40 years of age and for experience (low/high) with a cutoff score of 12 years teaching experience. Regarding the four types of dynamic assessment information, as rated in the follow up questionnaire, Mann Whitney U tests were conducted on each of these variables for age and teaching experience. The analyses resulted in a significant difference regarding *practiced problem solving strategies* for age:  $z = -2.16, p = .03$ , meaning that younger teachers valued information regarding the strategies of the children as more valuable than older teachers. A large majority of the younger teachers rated this information as a lot useful for constructing individual plans, whereas two-third of older teachers rated this as not at all or little useful. Evaluation of *improvement after hints* showed a significant difference for teaching experience:  $z = -2.00, p = .045$ , indicating that teachers with less experience rated this information as more positive than teachers with more experience. While all teachers with high experience valued this information regarding improvement after hints as 'not at all' or 'little useful', almost half of the teachers with less experience considered this as a lot useful.

Mann Whitney U tests were also conducted for age and experience regarding both the meaningfulness of the reports and recommendations, and the applicability of the recommendations. Only a trend was found for age on the meaningfulness of the reports ( $z = -1.90, p = .057$ ), which might indicate that younger teachers consider the reports and recommendations as more meaningful. A large majority of the younger teachers valued the reports above average, whereas only a quarter of the older teachers did so. Analyzing the possibility of the application of the recommendations by the teacher by a Mann Whitney U test did not reveal significant effects for age or teaching experience. However, inspection of crosstabs revealed different distributions than expected for both age ( $\chi^2 (3) = 8.01, p = .46$ ) and experience ( $\chi^2 (3) = 11.53, p = .009$ ). Older and more experienced teachers did value these recommendations as only 'a little' to 'reasonable' applicable; contrarily, half of the younger and less experienced teachers gave ratings as 'well applicable'. Age and experience of teachers clearly might have influenced our results.

## Discussion

In this study we examined the usability of static psychological reports versus reports based on dynamic assessment, to gain insight into teacher preferences and values regarding the meaningfulness of both types of assessment information. We aimed to investigate the supposed differences in teachers' appraisals with two versions of reporting. Children with severe math difficulties attending regular classrooms in which the teachers were working were subjects of these reports. In a former study we already investigated teachers' opinions of and recommendations based on reports written on outcomes regarding dynamic testing. Results were not always consistent, partly because the participating children were all typically developing and attended regular elementary schools. This first study therefore did not resemble the diagnostic process, whereby it was hypothesized that teachers could not express their evaluations regarding both the dynamic testing procedure and the subsequent recommendations (Bosma & Resing, 2008).

In the current study, therefore, our attention was focused upon the assessments and recommendations regarding elementary school children with severe learning difficulties within the math domain. We investigated which type of information teachers valued as useful for both classroom practice and formulating educational plans regarding individualized instruction.. Teachers' opinions were gathered by means of a (semi)diagnostic process within the regular school setting, including dynamic assessment, reports and recommendations for each participating child. To capture teachers' opinions and values regarding these reports and recommendations, for each participating child an individual report was written, either on static or on dynamic testing information. Further, the study investigated whether classroom practice and teacher estimations of learning potential were affected by the dynamic reports. A final aim was to capture teacher's opinion about the applicability of the provided recommendations through a follow up questionnaire.

*Dynamic testing reports.* As expected, at the end of the study the majority of teachers who received dynamic reports globally valued reports and recommendations as meaningful to very meaningful for constructing individual educational plans for the children.. However, almost equal appreciation was found in a majority of the values of teachers receiving static reports. We did observe only a few trends regarding differences in value of information for understanding cognitive functioning. Teachers given dynamic reports valued the reported observation and task behavior and the need for instruction slightly more positive than teachers who received static test reports. Presumably, our reports contained interesting, new information for a majority of teachers in both groups. Teachers appeared in particularly interested in recommendations regarding math difficulties, which we stressed in both report forms, although we intended to be more elaborated regarding the type and amount of needed instructions in the dynamic than in the static reports. Apart from the dynamic or static test results we presented teachers comparable information of results on math and memory tasks, observations, and task behavior during the tests, as a professional psychological report would have consist of.

One reason for not finding the expected differences between the teacher values after given them the different report forms could be that this baseline information might have been equally relevant and interesting for all teachers. As a consequence, it might have been the case that the surplus information regarding the need for instruction of a child, as provided in the dynamic testing reports, has not been visible or lengthy enough. Based on the two studies we did to date, we have to conclude that the value of dynamic assessment information in psychological reports of actual children in a need based context, appears to be very complex to tackle in a semi-diagnostic field-study.

*Teaching practice and learning potential.* It was also studied whether teaching practices would be differently affected by our intervention, the assessment and reports. We were interested whether in general teaching practice and interactions would change at all and between conditions. Classroom observations were conducted first at the start of the experiment and then repeated four weeks after the reports and recommendation were provided. It appeared that teachers, irrespectively the report form they had been given, at the second observation showed more intentional involvement with their



children, more frequent task regulating activities, more positive and informed feedback, and that they were observed to be better able to balance the different needs of the children during their general teaching, i.e., their math instruction. Although we did aim on differential results as a consequence of dynamic and static reporting and did not find these, it still can be concluded that reporting about children's math difficulties and the reported needs to solve the test tasks, including, for example, positive support, adapting the curriculum, somewhat changed the instructional practices of the majority of all teachers.

Specific interactions between the teacher and the participating child were observed as well. It was found that teachers receiving dynamic testing reports showed significantly more intentional involvement in interactions with their children. Regarding the use of praise and feedback we noticed a different change. Whereas teachers given static reports were observed to give more frequently positive feedback, the other teachers were observed to do so less frequently, if compared to the first observation. A similar pattern was found for challenging the ZPD. Teachers with the static reports, in contrast to our expectation, challenged their children significantly more than teachers with the dynamic reports, if compared to the first observation. These results cannot be explained by the report contents since these were kept similar, except for the dynamic testing results, and we further did not report the first observations to teachers. We have no other explanation for these findings.

Another question was related to teacher ratings and estimations of learning potential of the participating children. We expected that these ratings would be affected by the reports; in particular we expected that the second rated learning potential of children in the experimental group, after reading reports and recommendations, would relate stronger to the reported learning potential, compared to the first rating which took place at the start of the experiment. Results demonstrated that, as expected, teachers rated the learning potential of their children as below to low average at the beginning of our assessment, and rated it higher at the end of the experiment, but again we could not find differential effects. between the two teacher report groups.

In addition, we found for the teachers receiving the dynamic testing reports that the majority of their children were categorized as having an average to above average learning potential test outcome, despite of having math difficulties, while teachers rated them as below to low average. When these teachers however were asked to rate children's learning potential again, they judged them in higher learning potential categories that better corresponded with the actual reported learning potential. We therefore carefully conclude that on the rating of learning potential the reports had some positive influence, which corresponds to findings of, among others, Delclos et al. (1993), who concluded that teachers appeared to change their expectations about their students performance after reading reports based on dynamic assessment.

*Applicability of recommendations.* The follow up session, in which a sample dynamic report was provided to teachers, who had received static test reports and vice versa, appeared to be of interest. Teachers receiving the dynamic reports ascribed relatively higher values to the results and conclusion sections of the reports, compared to teachers receiving the static reports. For the recommendation sections teachers with the dynamic reports tended to rate the surplus value as significantly higher than teachers with the static reports. As this latter group of teachers read dynamic report information and compared this to the static reports they had received weeks earlier, this higher value was expected; the report contained additional information and in the recommendations the need for and type of instruction was taken into account. This finding, although a trend, gives us a glimpse of the potential value of dynamic assessment for the teaching practice. Further support we found in the ratings of the usefulness of dynamic assessment results for constructing individual plans. Teachers rated excerpts of a typical dynamic report, resembling Freeman and Miller's (2000) questionnaire, Although the variation between teachers was high, our findings indicate that a majority of teachers considered these specific types of dynamic assessment information as a reasonable to good contribution in the construction of individual educational plans. These outcomes tend to correspond to comparable ratings of special education coordinators in the study of Freeman and Miller (2000) who also recognized the usefulness of dynamic assessment information.

Results from a second interview regarding the meaningfulness of the reports showed that the majority of teachers receiving dynamic testing reports rated our recommendations as meaningful to very meaningful and appreciated the new, elaborated information regarding their children. Many teachers receiving the static reports, responded that, although the information was not completely new, it confirmed their general picture of the child's functioning and they therefore rated the meaningfulness of the recommendations as neutral to positive. We found similar results in the follow up questionnaire, showing that after some time teachers still valued the reports.

The follow-up data showed that younger and less experienced teachers valued the information regarding dynamic assessment results and recommendations relatively higher than their older and often more experienced peers. Younger teachers also tended to rate the overall value of the received reports and recommendations higher, and attached significantly higher values at the information concerning practiced strategies of the child. Teachers with lesser experience were observed to rate the value of information concerning 'improvement after hints' as significantly higher than their more experienced peers. Although these effects were not found in previous studies regarding preference for psychological reports or dynamic assessment (e.g., Delclos et al, 1993; Hulburt, 1995), seniority and experience may less positively influence the attributions towards dynamic assessment, possibly because they overrate their own impressions of the potential of the child (e.g., Richardson & Placier, 2001).

Although the overall results in our study showed that teachers recognized the (potential) value of dynamic assessment, we were not able to demonstrate differences in report preferences of teachers in the experimental or control condition. One reason for the absence of differences might lay in our design. Teachers in both conditions received reports about each of their children as if these children were referred to a professional educational psychologist. We therefore did not just list test results, but integrated these in the typical format of a results, conclusion and recommendation section. However, it appears that these report types including the reported information, were very new and relevant for teachers in both groups. Delclos et

al (1993) did not find report preference between teachers who got a dynamic versus a static report as well; they found however, that the background of teachers, especially their type of teacher training, did play a role in understanding the new type of reports. Teachers in our experiment all had finished the standard teacher education program and training and there were no differences between conditions for teaching experience, age and the experience with reading psychological reports. Teachers in the study of Delclos et al. (1993), who had been trained in mediated learning, had better understanding of the dynamic reports, compared to teachers trained in direct learning. It might therefore be necessary to inform teachers about the theory regarding dynamic assessment first so that they can acquire a framework for reading and understanding dynamic reports.

The fact that we chose to report the dynamic test results together with the static results of the math and memory tests, and task behaviors as observed during testing, which enabled us to provide appropriate and adequate recommendations, might have limited our results as we indicated earlier. A design with a control group in which teachers only receive pure test results, without interpretation, might lead to different findings regarding report preferences, but leads probably into difficulties in convincing and motivating teachers to participate.

Studies regarding interventions in the classroom (e.g., Witt, 1986) also acknowledged teachers' concerns related to amount of time and material resources that an intervention requires. Teachers appear to have a preference for interventions requiring as less time as possible. Besides time, it is known that teachers prefer practical recommendations (e.g., Haywood & Lidz, 2007), but in this study we also observed that for some teachers our practical recommendations were not concrete and practical enough. These teachers preferred recommendations with reference to materials and fully prescribed guidelines for implementation and organization in the classroom.

Many studies have mentioned and pointed out the advantages of dynamic assessment for diagnostic, educational and intervention purposes (e.g., Grigorenko, 2009; Haywood & Lidz, 2007). These advantages have not only been pointed out by researchers in the field of dynamic assessment, but were also noted by educational psychologists (Deutsch & Reynolds, 2000), special education coordinators (Freeman &

## *Chapter 6*

Miller, 2000), and others with expertise on assessment methods and applications (e.g., Fuchs et al., 2007; Pameijer, 2006). Teachers of in this study participating children with math difficulties acknowledged the potential advantages of dynamic assessment as we also found in our previous study (Bosma & Resing, 2008) regarding typical developing children. Teachers, especially older and more experienced, might not fully understand or appreciate the ideas, aims and outcomes of dynamic assessment, as we presented these in our reports. It therefore seems very necessary to focus more on dynamic assessment in teacher education than now is the case in most countries. Only if knowledge and understanding of principles of dynamic assessment will become more widespread under teachers and other professionals in education, the request for dynamic assessment will increase, which will show the use and advantage of dynamic assessment in providing concrete and practical recommendations for classroom interventions and the construction of educational plans.