Differentiation in primary mathematics education

Emilie Prast

Cover	Eva van de Weijer-Bergsma
Photographs	Eva van de Weijer-Bergsma, Eric van den Bandt (p.180), Margreet van Schie (p.186)
Printed by	ProefschriftMaken
Layout	Renate Siebes | Proefschrift.nu
ISBN	978-90-393-6983-8

Differentiation in primary mathematics education

Differentiatie in het primair rekenonderwijs

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht op gezag van de rector magnificus,
prof.dr. G.J. van der Zwaan, ingevolge het besluit van het
college voor promoties in het openbaar te verdedigen
op vrijdag 18 mei 2018 des ochtends te 10.30 uur

door

Emilie Johanna Prast

geboren op 31 december 1988
te Leiden

Promotoren: Prof.dr. J.E.H. van Luit
Prof.dr. E.H. Kroesbergen

Copromotor: Dr. E. van de Weijer-Bergsma

Dit proefschrift werd mede mogelijk gemaakt met financiële steun van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO; projectnummer 411-10-753).

Contents

1

General introduction

1.1 Differentiation

Students differ from each other in multiple ways – e.g., regarding intelligence, previous educational experiences, social economic status, and motivation – which all have an impact on what they have already learned and what they need to progress in their learning. Primary school classrooms are traditionally diverse since, in most countries, primary school has only one mainstream track for students of diverse achievement levels (in contrast to secondary school, in which students are often placed in separate tracks based on achievement level). Due to the current movement towards inclusion of children with special educational needs in general education classrooms, the range of academic ability and achievement levels within primary school classrooms is increasing even further. Thus, students within one classroom may have widely varying zones of proximal development (Vygotsky, 1978). Therefore, learning content that is appropriate for most students in the class may be too easy for some students and too difficult for others. Teaching students with different zones of proximal development poses a challenge for teachers: How to adapt education to the varying educational needs of students of diverse achievement levels? In other words: How to differentiate education?

Differentiation in primary mathematics education is the central theme of this dissertation. The focus is on differentiation based on students' current achievement level, also called cognitive or readiness-based differentiation. Differentiation is defined as 'an approach by which teaching is varied and adapted to match students' abilities using systematic procedures for academic progress monitoring and data-based decision-making.' (Roy, Guay, & Valois, 2013, p.1187). According to this definition, teachers should monitor students' academic progress to identify students' educational needs and then adapt instruction and practice to these needs. The way in which progress is monitored and the nature of educational adaptations can vary substantially, and various organisational formats can be used. One frequently used way to organise differentiation is homogeneous within-class ability grouping, in which students with similar academic ability or achievement levels are placed together in subgroups within the heterogeneous classroom (Tieso, 2003).

1.2 Adaptive teaching competency

For teachers, implementing differentiation requires advanced knowledge and skills. In line with Vogt and Rogalla (2009), the term 'adaptive teaching competency' is used in this dissertation to refer to teachers' capacities for making adaptations to students' identified educational needs (i.e., implementing differentiation). Adaptive teaching competency is defined in terms of four dimensions: subject matter knowledge, the ability to diagnose students' current understanding and achievement, the ability to use diverse teaching methods to meet diverse students' needs, and classroom management skills. Thus, it requires both general pedagogical skills (e.g., classroom management) and domain-specific subject matter knowledge (e.g., in mathematics) as well as pedagogical content knowledge (e.g., didactical models). The definition of adaptive teaching competency implies that differentiation is strongly grounded in a particular content domain. In the domain of mathematics, for example, teachers should not only master the content which is to be learned by the students themselves, but also possess pedagogical content knowledge regarding (a) how mathematics is typically learned (e.g., how students come to understand mathematical concepts), (b) how mathematics is typically taught (e.g., the order in which specific solution strategies are taught), (c) how specific educational needs can be diagnosed (e.g., why does a student struggle with a particular type of sums?), and (d) how specific educational needs can be met (e.g., how to adapt instruction when a student does not understand the concept of multiplication). In order to study differentiation in sufficient depth and to enhance the potential applicability of the findings in teachers' daily practice, this dissertation zooms in on one content domain, namely primary mathematics education.

1.3 Primary mathematics education in the Netherlands: Need for differentiation

Mathematics is one of the core subjects in primary school and a basic understanding of mathematics is necessary to function in society. In international comparisons of mathematics achievement, Dutch students are losing their traditionally high place in the rank order. This is not only due to increased mathematical competence in other countries, but also to a slow but steady decrease in mathematical competence in the Netherlands over the last twenty years (Meelissen & Punter, 2016). Moreover, while relatively many Dutch students reach at least a basic achievement level, only few Dutch students reach an excellent achievement level compared to students in other countries (Meelissen & Punter, 2016). These findings are a cause for concern and have been linked to teachers' potentially insufficient competence in mathematics (i.e., their own skill level), didactics of mathematics, and differentiation in mathematics (KNAW, 2012). Accordingly, a need for additional training (for pre-service teachers) and professional development (for in-service teachers) about primary mathematics education in general, and differentiation in particular, has been identified (Inspectie van het Onderwijs, 2012, 2015; KNAW, 2012; Schram, Van der Meer, & Van Os, 2013).

1.4 Project GROW

1.4.1 Goals

Against this background, project GROW ('Gedifferentieerd RekenOnderWijs', i.e., a Dutch acronym for differentiated primary mathematics education) was launched to study differentiation for students of diverse achievement levels in primary mathematics education in more depth and, ultimately, to develop and evaluate a teacher professional development (PD) programme about this topic[1]. The first goal of the project was to specify what differentiation entails in the context of primary mathematics education. The term 'differentiation' is very broad and has been used in multiple ways. Moreover, strategies for differentiation are strongly grounded in the content domain in which they are to be applied. Thus, there was a need to specify what teachers should do in order to meet the needs of students with diverse achievement levels in mathematics. The second goal was to investigate the degree to which teachers already implement differentiation (before PD about this topic). Only a few previous studies had examined the implementation of differentiation for students of a broad range of achievement levels (e.g., Roy et al., 2013) and there was still a need for studies specifically in the domain of mathematics as well as observational studies. Besides the overall level of implementation, it was examined whether certain strategies for differentiation were used relatively frequently or infrequently. This could inform teacher educators about the relative ease or difficulty of implementing specific strategies and provide directions regarding areas in which there is most room for improvement. The third goal was to develop, implement and evaluate a professional development (PD) programme about differentiation in mathematics. The PD programme was aimed at meeting the educational needs students of *all* achievement levels, i.e., including low-achieving, average-achieving, and high-achieving students. Although (PD about) differentiation is widely recommended, little is known about its effects on teachers' instructional behaviour or on students' achievement. While previous studies have demonstrated positive effects of two technological applications for differentiation (reviewed by Deunk, Doolaard, Smale-Jacobse, & Bosker, 2015), studies in which the teacher has a central role in implementing differentiation are scarce. Specifically, there is still a need for large-scale studies investigating the effects of PD on teachers' observable behaviour and student achievement. The fourth goal was to investigate how students' achievement level is reciprocally related to motivation for mathematics in students of diverse achievement levels. Previous achievement is theorised to be an important source of aspects of motivation such as self-efficacy (Bandura, 1997) and self-concept (Marsh & Martin, 2011). In turn, motivation is supposed to promote adaptive learning behaviours

[1] Project GROW was funded by the Netherlands Organisation for Scientific Research (NWO; grant number 411-10-753).

such as persistence, which should have a positive effect on future achievement (Marsh & Martin, 2011; Wigfield & Eccles, 2000). Since the educational experiences of low-achieving, average-achieving, and high-achieving students are likely to differ substantially – for example, low-achieving students may experience failure more often – the relations between motivation and achievement may also differ depending on achievement level. Therefore, the relations between motivation and achievement were investigated over time with particular attention for whether and how these relations differed between low-achieving, average-achieving and high-achieving students. This could provide new knowledge about the relative importance of several aspects of motivation for students of diverse achievement levels, which could have implications for desirable differentiation practices.

1.4.2 Design

In project GROW, researchers of Utrecht University collaborated intensively with a consortium of pre-service and in-service teacher educators with expertise in differentiation and primary mathematics education. This collaboration was sought in order to enhance the compatibility between research and the daily practice of teaching and teacher education.

An overview of the project is presented in Figure 1.1. In the first phase of the project, the consortium members participated in an expert consensus procedure to specify what differentiation in primary mathematics education entails. A combination of focus group discussions and the Delphi method was used to achieve consensus among the consortium members on a model and recommended strategies for differentiation. Based on the specification of differentiation resulting from the consensus procedure, a PD programme about differentiation in primary mathematics education was developed in continued collaboration with the consortium of experts.

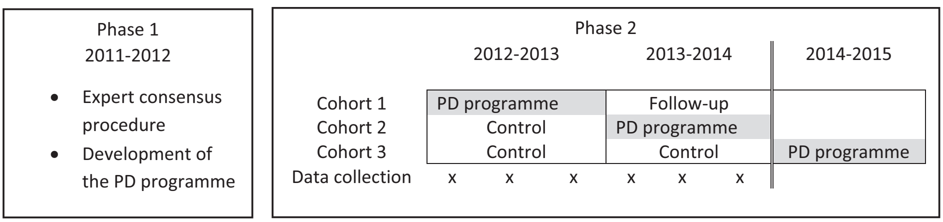


Figure 1.1 Design of project GROW. PD = Professional development

In the second phase of the project, the PD programme was implemented and evaluated in a large-scale study involving 32 whole primary schools (*N* = 400 teachers and 6187 students of grade 1 through 6). Participating schools were randomly assigned to one of three cohorts. In each cohort, data were collected across two schoolyears (i.e., all schools provided data on all measurement occasions), but the timing of the intervention differed between the cohorts: Cohort 1 participated in the PD programme in Year 1 and was a follow-up condition in Year 2, Cohort 2 was a control condition in Year 1 and participated in the PD programme in Year 2, and Cohort 3 served as a control condition in both years. Data were collected at the beginning, middle, and the end of each of the two schoolyears. In the 2014-2015 schoolyear, the evaluation study had ended but schools of Cohort 3 could participate in the PD in return for their participation in data collection during the two previous years.

1.4.3 Teacher-level measures

At the teacher level, both self-report and observational data were collected to measure teachers' implementation of differentiation. No instruments were available to measure the application of differentiation in the domain of mathematics in sufficient detail for our purposes. Therefore, two new instruments were developed and administered: a self-report instrument called the Differentiation Self-Assessment Questionnaire (DSAQ) and a video observation instrument called the Differentiation in Mathematical Instruction (DMI). At the beginning of both school years, teachers were asked to report their own implementation of a range of strategies for differentiation using the DSAQ. In addition, video observations of mathematics lessons were carried out at the beginning and end of both school years in a subsample of teachers. The collected videos were scored with the newly developed DMI, a video observation instrument to score teachers' implementation of differentiation in mathematics.

1.4.4 Student-level measures

At the student level, achievement in mathematics was measured using the standardised, nationally administered Cito Mathematics Test (CMT; Janssen, Scheltens, & Kraemer, 2005). The CMT was administered at the middle and end of both schoolyears. Additionally, the scores of the end of the previous school year (June 2012) were used as the baseline measure, yielding a total of five timepoints at which mathematics achievement was measured. Student motivation for mathematics – including self-efficacy, self-concept, task value, and mathematics anxiety – was measured using a questionnaire which was newly developed for this purpose (since no previous questionnaire covered all these aspects of motivation in a format suitable for administration for students from grade two through six). Three additional variables which were known to affect mathematics achievement were administered to control for these variables: nonverbal intelligence, visual-spatial working memory and verbal working memory. Nonverbal intelligence was measured with the well-established Raven's Standard Progressive Matrices (SPM; Raven, Court, & Raven, 1996). Working memory was measured using two newly developed measures for online, self-reliant administration: the Lion game (Van de Weijer-Bergsma, Kroesbergen, Prast, & Van Luit, 2016) and the Monkey game (Van de Weijer-Bergsma, Kroesbergen, Jolani, & Van Luit, 2016).

1.5 This dissertation

An overview of the topics and structure of this dissertation is provided in Figure 1.2. Chapter 2 reports about two studies. Study 1 describes the method and results of the expert consensus procedure which was used to specify what differentiation entails. Study 2 presents the newly developed Differentiation Self-Assessment Questionnaire and describes how teachers rated their own usage of the differentiation strategies recommended by the experts in Study 1. Chapter 3 reports about the results of the video observations which were scored with the newly developed Differentiation in Mathematical Instruction. First, the chapter describes teachers' observed implementation of differentiation – including relatively frequently and infrequently used strategies – at baseline. Second, the chapter analyses the effects of the PD programme on teachers' observed implementation of differentiation. Chapter 4 investigates the effects of the PD programme on student achievement. Latent growth models were used to evaluate the short-term and long-term effects on mathematics achievement growth after controlling for nonverbal intelligence and working memory. Chapter 5 investigates the longitudinal and potentially reciprocal relations between achievement and several aspects of motivation for mathematics (self-efficacy, self-concept, task value, and mathematics anxiety). First, a mediation model in which the motivational variables were modelled as mediators between previous and subsequent achievement was developed in the total sample. Second, multiple-group

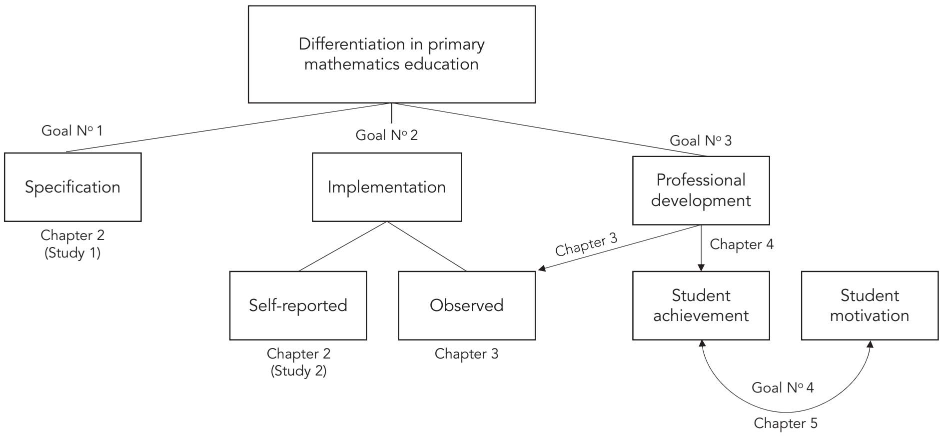


Figure 1.2 Overview of the topics and structure of this dissertation.

modelling was used to investigate whether the relations between achievement and motivation (as specified in the final model) were similar or different for students of diverse achievement levels. Finally, Chapter 6 provides a general summary and discussion of all chapters in this dissertation.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Deunk, M., Doolaard, S., Smale-Jacobse, A., & Bosker, R. J. (2015). *Differentiation within and across classrooms: A systematic review of studies into the cognitive effects of differentiation practices*. Groningen, The Netherlands: GION.

Inspectie van het Onderwijs. (2012). *De staat van het onderwijs: Onderwijsverslag 2010-2011* [The state of education: Educational report 2010-2011]. Utrecht, The Netherlands: Inspectie van het Onderwijs.

Inspectie van het Onderwijs. (2015). *Beginnende leraren kijken terug* [Beginning teachers look back]. Utrecht, The Netherlands: Inspectie van het Onderwijs.

Janssen, J., Scheltens, F., & Kraemer, J. M. (2005). *Rekenen-wiskunde groep 3–8: Handleidingen* [Mathematics test grade 1 through 6: Manuals]. Arnhem, The Netherlands: Cito.

KNAW. (2009). *Rekenonderwijs op de basisschool: Analyse en sleutels tot verbetering* [Mathematics education in primary school: Analysis and keys to improvement]. Amsterdam, The Netherlands: KNAW.

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*, 59-77. doi: 10.1348/000709910X503501

Meelissen, M. R. M., & Punter, R. A. (2016). *Twintig jaar TIMSS: Ontwikkelingen in leerlingprestaties in de exacte vakken in het basisonderwijs 1995 - 2015* [Twenty years of TIMSS: Developments in student achievement in mathematics and science in primary education 1995 - 2015]. Enschede, The Netherlands: Universiteit Twente.

Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's standard progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press.

Roy, A., Guay, F., & Valois, P. (2013). Teaching to address diverse learning needs: Development and validation of a differentiated instruction scale. *International Journal of Inclusive Education*, *17*, 1186-1204. doi:10.1080/13603116.2012.743604

Schram, E., Van der Meer, F., & Van Os, S. (2013). *Omgaan met verschillen: (G)een kwestie van maatwerk* [Responding to differences: (Not) a matter of customisation]. Enschede, the Netherlands: SLO.

Tieso, C. L. (2003). Ability grouping is not just tracking anymore. *Roeper Review*, *26*, 29-36. doi: 10.1080/02783190309554236

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Jolani, S., & Van Luit, J. E. H. (2016). The monkey game: A computerized verbal working memory task for self-reliant administration in primary school children. *Behavior Research Methods*, *48*, 756-771. doi:10.3758/s13428-015-0607-y

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Prast, E. J., & Van Luit, J. E. H. (2015). Validity and reliability of an online working memory task for self-reliant administration in school-aged children. *Behavior Research Methods*, *47*, 708-719. doi:10.3758/s13428-014-0469-8

Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, *25*, 1051-1060. doi:10.1016/j.tate.2009.04.002

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68-81. doi:10.1006/ceps.1999.1015

2

Readiness-based differentiation in primary school mathematics: Expert recommendations and teacher self-assessment

Emilie J. Prast
Eva Van de Weijer-Bergsma
Evelyn H. Kroesbergen
Johannes E. H. Van Luit

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2015). Differentiation in primary school mathematics: Expert recommendations and teacher self-assessment. *Frontline Learning Research*, *3*, 90-116. doi:10.14786/flr.v3i2.163

Author contributions: All authors designed research; EvdW and EP collected the data under supervision of HvL and EK; EP analysed the data and wrote the paper; EvdW, EK and HvL provided feedback on the paper.

Abstract

The diversity of students' achievement levels within classrooms has made it essential for teachers to adapt their lessons to the varying educational needs of their students ('differentiation'). However, the term differentiation has been interpreted in diverse ways and there is a need to specify what effective differentiation entails. Previous reports of low to moderate application of differentiation underscore the importance of practical guidelines for implementing differentiation. In two studies, we investigated how teachers should differentiate according to experts, as well as the degree to which teachers already apply the recommended strategies. Study 1 employed the Delphi technique and focus group discussions to achieve consensus among eleven mathematics experts regarding a feasible model for differentiation in primary mathematics. The experts agreed on a five-step cycle of differentiation: (1) identification of educational needs, (2) differentiated goals, (3) differentiated instruction, (4) differentiated practice, and (5) evaluation of progress and process. For each step, strategies were specified. In Study 2, the Differentiation Self-Assessment Questionnaire (DSAQ) was developed to investigate how teachers self-assess their use of the strategies recommended by the experts. While teachers (N = 268) were moderately positive about their application of the strategies overall, we also identified areas of relatively low usage (including differentiation for high-achieving students) which require attention in teacher professional development. Together, these two studies provide a model and strategies for differentiation in primary mathematics based on expert consensus, the DSAQ which can be employed in future studies, and insights into teachers' self-assessed application of specific aspects of differentiation.

2.1 Introduction

Every day, primary school teachers are faced with the task of teaching students of diverse academic ability and achievement levels. Therefore, teachers should adapt their lessons to the diverse educational needs of their students (Corno, 2008). Such adaptations are often promoted using the term *differentiation* or *differentiated instruction*, defined by Tomlinson et al. (2003, p.120) as "an approach to teaching in which teachers proactively modify curricula, teaching methods, resources, learning activities, and student products to address the diverse needs of individual students and small groups of students to maximise the learning opportunity for each student in a classroom".

The international trend towards inclusive education makes the need for differentiation especially urgent. Within response to intervention models, general education teachers are required to provide both universal support – i.e., a good general education for all students (Tier 1) – and targeted support (Tier 2) such as small-group instruction for struggling students (Fuchs & Fuchs, 2007; McLeskey & Waldron, 2011). Small-group or individual interventions carried out by an educational specialist (Tier 3) are only available for a limited number of students whose problems persist despite the supports provided by the general education teacher. Thus, general education teachers have the primary responsibility for providing a good education to all students, regardless of their achievement level.

Attending to the educational needs of students with a broad range of ability and achievement levels is a challenge for teachers. Successful differentiation requires advanced subject matter knowledge, pedagogical skills and classroom management skills (VanTassel-Baska & Stambaugh, 2005). Consequently, a need for professional development in the area of differentiation has been identified repeatedly (Johnsen, Haensly, Ryser, & Ford, 2002; Inspectie van het Onderwijs, 2012; VanTassel-Baska et al., 2008).

To design effective professional development programmes, it is important to know what teachers should do in their day-to-day teaching to differentiate their lessons for students of diverse achievement levels. What constitutes best practice? In two studies, we investigated how teachers should differentiate according to experts as well as the degree to which teachers already apply the recommended strategies. The focus was exclusively on mathematics since strategies for differentiation may vary across subject areas. Moreover, domain-specific guidelines or strategies tend to be more concrete and may therefore provide stronger guidance to teachers.

Differentiation is an umbrella term that may be used to refer to one or several of a variety of instructional modifications. It may involve modifications of the content (what students learn), the process (how they learn it), or the product of learning (how students demonstrate their learning) (Tomlinson, 2005). Various student characteristics may serve as a ground for differentiation. For example, Tomlinson et al. (2003) distinguish between differentiation by student readiness (representing the current level of knowledge and skills in the subject area), learning profile (a student's preferred ways of learning, such as a preference for visual input) and interest (topics about which the student wants to learn more).

In the current study, the focus is on differentiation by student readiness. Readiness is influenced by a child's natural ability as well as learning experiences and is reflected in the child's current knowledge and skill level. The importance of differentiation by student readiness is supported by the theoretical constructs of the zone of proximal development (Vygotsky, 1978), challenge-skill balance (Csikszentmihalyi, 1990), aptitude-treatment interaction (Cronbach & Snow, 1977), and adaptive teaching (Corno, 2008). Vygotsky (1978) stated that learning occurs when a child engages in activities that fall within its zone of proximal development (ZPD), i.e. that are slightly more difficult than what the child already masters independently. When children within one classroom have widely varying readiness levels, their zones of proximal development also differ. A task that is just within reach for average-achieving students (i.e. in their ZPD) may be too difficult for children with lower readiness levels when the gap between existing knowledge and skills and the task is too big. Conversely, children with higher readiness levels may already master the task and in this case they are not challenged to reach beyond what they can already do. This implies that children within the same classroom may need different instructional treatments to work in their ZPD. To work in the ZPD, the skill level of the students should be in balance with the difficulty level of the tasks. Such a challenge-skill balance may result in effective and engaged learning, while tasks that are much too difficult or too easy may lead to frustration, boredom, and withdrawal from learning (Csikszentmihalyi, 1990). Additionally, certain characteristics of the learning environment may be useful for some learners but not for others, depending on the aptitude of the student (Cronbach & Snow, 1977). Because of the variation in student aptitudes and the resulting diversity of educational needs, teachers should adapt education to the needs of their students (Corno, 2008). What these theories have in common is the idea that students with different readiness levels have different educational needs and that instruction should be matched to these needs, which is exactly what differentiation aims to do.

Roy, Guay, and Valois (2013) took a step towards clarification of the term differentiation by identifying two main components of readiness-based differentiation: academic progress monitoring and instructional adaptations. Ideally, the developments in students' achievement or understanding are closely followed, for example using frequent formal or informal tests, and adaptations are then made to ensure a good fit between the readiness of the student and the instruction.

Most approaches to differentiation include these two components in some way. Nevertheless, the way in which progress is monitored and the nature of instructional adaptations strongly vary across intervention studies (e.g. Brown & Morris, 2005; McDonald

Connor et al., 2009; Reis, McCoach, Little, Muller, & Kaniskan, 2011; Tieso, 2005; Ysseldyke & Tardrew, 2007). Students' achievement may be measured with standardised, curriculum-based, or informal assessments. In some cases, the results of these assessments are used to determine the instructional treatment for an extended period of time (weeks or months) whereas other interventions continuously monitor progress and adapt the instructional treatment accordingly. Adaptations may be at the level of individual students or subgroups of students. When grouping is used, such groups may be between-class or within-class, fixed or flexible (Tieso, 2003). Adaptations may entail modification of the amount of instruction, the content or type of instruction, the content or type of independent practice tasks, or combinations of these elements. Given the diverse interpretations of the term differentiation, there is a need to specify what effective differentiation entails.

One line of research has examined the effects of various types of ability grouping. The best results are obtained when students can switch between groups based on changes in their educational needs (the progress monitoring component of differentiation) and when instruction is tailored to the needs of the students in the groups (the instructional adaptations component) (Kulik & Kulik, 1992; Lou et al., 1996; Slavin, 1987; Tieso, 2003). When these conditions are met, homogeneous within-class ability grouping has demonstrated positive effects on student achievement across multiple studies (Kulik & Kulik, 1992; Lou et al., 1996; Slavin, 1987; Tieso, 2005). In contrast, slight negative effects of within-class ability grouping in primary school were found across three studies in which variations in instructional treatment were not explicitly described (Deunk, Doolaard, Smale-Jacobse, & Bosker, 2015). So, it seems to be important to use the grouping arrangement as a means to provide the different subgroups with the instruction that they specifically need, i.e. to differentiate instruction. Another issue in the literature on ability grouping is the potential existence of differential effects, i.e. different effects for students of different ability levels. While Slavin (1987) reported a higher median effect size for low-ability students than for average-ability and high-ability students, other reviews have found different patterns with smaller or even negative effects for low-achieving students (Deunk et al., 2015; Kulik & Kulik, 1992; Lou et al., 1996). More research is necessary to determine in which situations such differential effects may arise.

A recent review (Deunk et al., 2015) examined the effects of various readiness-based differentiation practices on student achievement. Although the authors aimed to include all high-quality studies published about this topic since 1995, only sixteen studies about differentiation in primary school could be included. Most of these sixteen studies were either too narrow (ability grouping without explicit instructional differentiation) or too broad (interventions in which differentiation was one of several components of a comprehensive school reform initiative) to be informative about the effects of differentiation on student achievement. However, promising results were obtained with two computerised interventions for differentiation: Individualizing Student Instruction (McDonald Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007; McDonald Connor et al., 2011a; McDonald Connor et al., 2011b) and Accelerated Math (Ysseldyke et al., 2003; Ysseldyke & Bolt, 2007). The Individualizing Student Instruction programme provides the teacher with recommendations about the amount and type of literacy instruction needed by individual students based on their scores on a computerised test. Accelerated Math is a technological application which continuously monitors students' progress, adapts practice tasks to students' individual skill level, and informs the teacher when students struggle with certain types of problems. Both of these interventions, which clearly contain both components of differentiation (progress monitoring and instructional adaptations), have demonstrated significant positive effects across multiple studies.

Prior research has shown that there is room for improvement in teachers' implementation of differentiation. The Dutch Inspectorate of Education recently found that adequate adaptations to diverse educational needs are only made at about half of the schools (Inspectie van het Onderwijs, 2012). In US middle schools, both teachers and students reported low usage of differentiation strategies (Moon, Callahan, Tomlinson, & Miller, 2002). In a recent study on Canadian elementary schools, teachers self-reported moderate use of differentiation strategies, but strategies requiring more time to implement were used relatively infrequently (Roy et al., 2013). Similarly, studies about adaptations for students with learning disabilities found that teachers tend to implement *typical* adaptations which can be easily implemented for all students rather than *specialised* adaptations, i.e. adaptations targeted at the unique educational needs of individual students (McLeskey & Waldron, 2002, 2011; Scott, Vitale, & Masten, 1998). However, a recent study carried out in Finland found that teachers do provide more individual support to struggling students (Nurmi et al., 2013). For high-achieving or gifted students, low levels of differentiation have generally been found (Reis et al., 2004; Westberg, Archambault, Dobyns, & Salvin, 1993; Westberg & Daoust, 2003). In sum, prior studies have generally found low to moderate use of differentiation strategies, although the degree of implementation of differentiation seems to vary depending on the specific strategies for differentiation examined, the targeted population of students, and perhaps also the country in which data are collected. Specialised adaptations as well as adaptations targeted specifically at high-achieving students seem to be used relatively infrequently.

In conclusion, there is a clear need to apply differentiation based on differences in students' readiness and teachers could use some help in doing this. The literature shows that differentiation should include progress monitoring and instructional adaptations. However, the ways in which this can be done effectively are less clear. Promising results have been obtained with two computerised interventions. However, high-quality research about the achievement effects of interventions in which differentiation is mainly implemented by the teacher himself is scarce. There is a need for general guidelines for differentiation that can be applied in a wide array of schools, independently from curricular methods or technological applications. Therefore, Study 1 sought to achieve consensus among a consortium of mathematics experts about a feasible model and associated strategies for differentiation. Study 2 linked the results of Study 1 to teachers' daily practice by examining how teachers self-assess their use of the strategies for differentiation recommended by the experts.

2.2 Study 1

2.2.1 Aims Study 1

The aim of Study 1 was to operationalise the concept of differentiation by achieving consensus among a consortium of mathematics experts about a coherent set of strategies for differentiating primary school mathematics education. The result of the consensus procedure needed to be feasible for use by general education teachers in daily mathematics teaching. Additionally, it needed to be applicable in diverse schools, independent from curricular method.

Expert consensus procedures can be valuable when scientific literature provides insufficient information to make complex decisions (Landeta, 2006) and have been applied before to achieve consensus about effective teaching (Teddlie, Creemers, Kyriakides, Muijs, & Yu, 2006). While several individual experts have made recommendations for differentiation in primary mathematics in books and journals for practitioners, consensus among various experts could provide a more solid foundation. For differentiation in mathematics, teacher educators with expertise in the didactics of primary mathematics are the relevant group of experts. Teacher educators may have gained practical knowledge regarding the effectivity and feasibility of diverse strategies for differentiation. Making use of this experiential knowledge has the potential to complement the scientific literature and strengthen the link between theory and practice.

2.2.2 Method Study 1

2.2.2.1 Participants

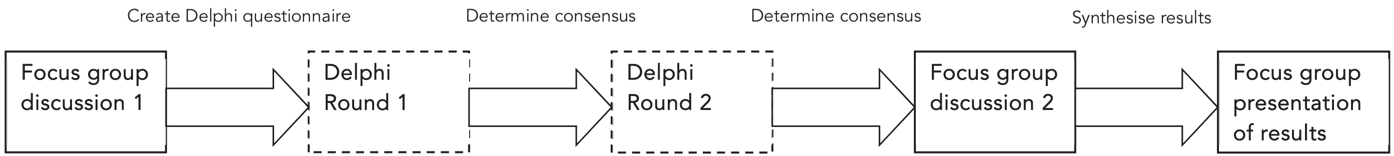
The consortium of experts was designed to include distinguished Dutch pre-service and in-service teacher educators with a professional focus on mathematics education. To be eligible for participation, potential members had to be experts in their field, as demonstrated by their (1) experience in providing pre-service or in-service teacher training about teaching mathematics (2) regular presence as invited speaker at educational conferences and (3) role as a consultant to the Ministry of Education, Culture and Science to discuss new educational policy. The senior authors approached potential candidates with these criteria in mind. All experts who were invited to participate agreed to join the consortium.

This resulted in a consortium of eleven experts (seven men, four women) representing eight large national and regional institutes for pre- and in-service teacher training spread across the Netherlands. The members had experience in at least two of the following areas: in-service teacher training for mathematics (*M* = 8.6 years, *SD* = 8.5 years), pre-service teacher training for mathematics (*M* = 5.4 years, *SD* = 6.3 years), carrying out educational evaluation studies (*M* = 25.0 years, *SD* = 21.2 years) and teaching (*M* = 5.7 years, *SD* = 5.4 years). The current daily work of the consortium members included educating pre-service teachers, providing professional development for in-service teachers, and guiding schools in the implementation of new educational approaches including differentiation.

2.2.2.2 Consensus procedure

Focus group discussions (Liamputtong, 2011) and the Delphi method (Hasson, Keeney, & McKenna, 2000) were used to investigate the experiential knowledge of the experts on differentiated mathematics education systematically. Focus group discussions are structured discussions with a group of persons involved in the topic in which certain roles (e.g. a discussion leader, a timekeeper and a secretary) and rules (e.g. only on-topic contributions) are specified and adhered to. The Delphi technique entails the repeated administration of a questionnaire in order to achieve consensus among experts. After the first round of administration, the initial responses are presented anonymously to the participants, who are then asked to fill out the questionnaire again. This procedure is repeated until consensus (specified with a consensus criterion) is reached. The order of focus group discussions and Delphi rounds in the current study is presented in Figure 2.1. The whole procedure took place between November 2011 and January 2012.

In the first three-hour focus group discussion, the experts were invited to share their knowledge, prompted by eight core questions about differentiation (see Figure 2.1). These questions were deliberately left open to elicit broad input. No particular theoretical perspective was chosen a priori apart from the assumption that student readiness would be an important ground for differentiation (see questions 6 and 7). Rather, the questions were asked from a practical point of view (what does and does not work in practice and how can this be improved). In principle, the questions were discussed one by one in the listed order, but in practice, the discussion sometimes moved back and forth between the various questions because of their high interrelatedness. After the first focus group discussion, the first author restructured the meeting minutes in terms of (initial) answers to the eight core questions.



<u>Questions asked in focus group discussion 1:</u>

1. What should differentiation in primary maths education look like?
2. What could be improved compared to the current situation on primary schools regarding differentiation in maths?
3. What does a teacher who successfully differentiates education do?
4. What are challenges in implementing differentiation in maths education?
5. Which types of instruction could be distinguished and for which pupils would these types of instruction be appropriate?
6. Are there other aspects than readiness level based on which a teacher could differentiate (for example: visual or verbal instruction)?
7. What should a teacher do differently with students of different readiness levels?
8. How often should additional or specific types of instruction be provided?

In <u>focus group discussion 2,</u> issues on which consensus had not yet been reached were discussed and resolved.

<u>Content of the Delphi questionnaires</u>

Themes and consensus criteria are reported in Table 2.1

Round 1 questionnaire:

- Summarising concepts (rated on five-point scale)
- Specific statements (rated on five-point scale)
- Comments (open answer)

Round 2 questionnaire:

- Round 1 concepts and statements on which consensus had not been reached in Round 1
- Comments provided by the participants in Round 1 included as new specific statements

Figure 2.1 Consensus procedure.

Second, based on this input, the researchers constructed an online Delphi questionnaire (Round 1). During the first focus group discussion, one of the experts had listed five general themes that are central to differentiation: organisation, goals, instruction, practice, and learning styles. These themes were used to structure the Delphi questionnaire. The theme 'differentiation in kindergarten' was added to account for aspects of differentiation specific to kindergarten (kindergarten is integrated in the Dutch primary school system). For each theme, the first author summarised the main ideas of the focus group discussion and proposed this to the other authors. Apart from some minor changes, the other authors agreed that these summaries accurately reflected what had been said in the discussion. These summaries were included in the Delphi questionnaire as one-paragraph concepts for differentiation (see Appendix 2.1). For each theme, statements about specific elements of the concept were also included (for example: 'The low-achieving subgroup profits from extended instruction'). The experts rated their agreement with the concepts and with the specific statements on a Likert scale ranging from 1 (*do not agree at all*) to 5 (*fully agree*). Additionally, open questions prompted participants to provide any comments they had.

Third, a Round 2 Delphi questionnaire was developed which included only those questions on which no overwhelming consensus had been reached in Round 1. The consensus criterion for Round 1 was that all responses should be at one end of the scale (i.e. either 4 and 5 or 1 and 2), with a maximum of one neutral response (3). The questions on which consensus had not yet been reached were presented to the participants again accompanied by a bar chart of the responses in Round 1. Additionally, comments provided by the participants in Round 1 were included as new questions in Round 2. Using a more lenient consensus criterion of maximally three neutral responses and the rest at one end of the scale, the researchers determined for which items consensus was achieved in Round 2.

Fourth, the researchers presented the results of the Delphi questionnaire to the experts during the second focus group discussion which lasted two hours. The items on which consensus had not been achieved were discussed to clarify misunderstandings (especially about the open comments provided by the participants in Round 1) and resolve conflicting opinions.

Fifth, the first author reviewed the meeting minutes of the focus group discussions and the responses to the Delphi questionnaire to synthesise the input, resulting in a proposed model for implementing differentiation. The other authors agreed with this model.

Sixth, the proposed model for differentiation was sent to all consortium members and discussed during a third one-hour meeting of the focus group.

2.2.2.3 Attendance rates of consortium members

Of the eleven consultants, six (54.5%) attended the first two focus group discussions and completed the two Delphi questionnaires and four (36.4%) completed three out of four components (i.e. either both discussions and one questionnaire or both questionnaires and one discussion). One participant only completed the Delphi questionnaire, after being informed about the content of the first focus group discussion in a separate meeting with the researchers. All members received the proposed model for differentiation by email and were given the opportunity to send any comments or questions, and six participants (54.5%) attended the third meeting in which the cycle was discussed.

2.2.3 Results Study 1

In Round 1 of the Delphi questionnaire, the experts agreed with the concepts for differentiation in instruction, differentiated goals, differentiated practice, differentiation based on learning styles, and differentiation in kindergarten which had been formulated based on the first focus group discussion. Positive consensus on the remaining concept (organisation of differentiation) was reached in Round 2. Table 2.1 provides an overview of the degree of consensus achieved on the specific statements in the two rounds of the Delphi questionnaire. Consensus was reached on 35 items in the first round and on an additional 25 original items in the second round, amounting to consensus on 74.1% of original items after two rounds. Regarding the new statements that were derived from the open comments in Round 1, consensus was reached on 46.0% of these statements in Round 2. The items on which consensus had not been reached after the second Delphi questionnaire were discussed in the second focus group discussion. Differences in interpretation of certain items were resolved and consensus was reached about the main issues. Items on which no consensus had been reached in the Delphi questionnaires often concerned issues about which the experts were unsure or had no pronounced opinion, including the importance of specific elements (e.g. videotaped instruction, mind maps, games, student choice) and preference for certain grouping formats (e.g. pairs or small groups). In the second focus group discussion, the overall conclusion about these elements and formats was that they all have their merits and that the choice is dependent on the situation, but that they are not crucial for differentiation.

The experts approved the model for differentiation which was created based on their input. The model, dubbed the *cycle of differentiation*, consists of the following five steps: identification of educational needs, differentiated goals, differentiated instruction, differentiated practice, and evaluation of progress and process (see Figure 2.2). A distinction is made between instruction and practice. Instruction refers to moments during which the teacher provides instruction to the whole class, subgroups of students, or individual students, whereas practice refers to moments during which students work on tasks, individually or in groups. These two can happen simultaneously, for example when the teacher provides instruction to a subgroup while other students are working on practice tasks. In the following paragraphs, we describe the key recommendations for each step in the cycle of differentiation provided by the experts in the consensus procedure.

Table 2.1 Overview of consensus on statements in the Delphi questionnaire

Theme Subtheme	Original statements			New statements in Round 2	
	No. of statements	Consensus after Round 1	Consensus after Round 2[a]	No. of statements	Consensus after Round 2
Organisation of differentiation	11	4	7	14	2
Differentiation in instruction					
General	5	3	4	14	4
Whole-class instruction	9	2	6	10	6
Subgroup instruction for low-achieving students	10	3	8	7	3
Subgroup instruction for high-achieving students	12	6	8	4	3
Differentiated goals	10	4	8	23	10
Differentiated practice	14	5	9	6	3
Differentiation based on learning styles	3	1	3	25	15
Differentiation in kindergarten	7	7	7	10	6
Total	81	35	60	113	52

[a] Total amount of items on which consensus was reached, including items on which consensus had already been reached in Round 1.

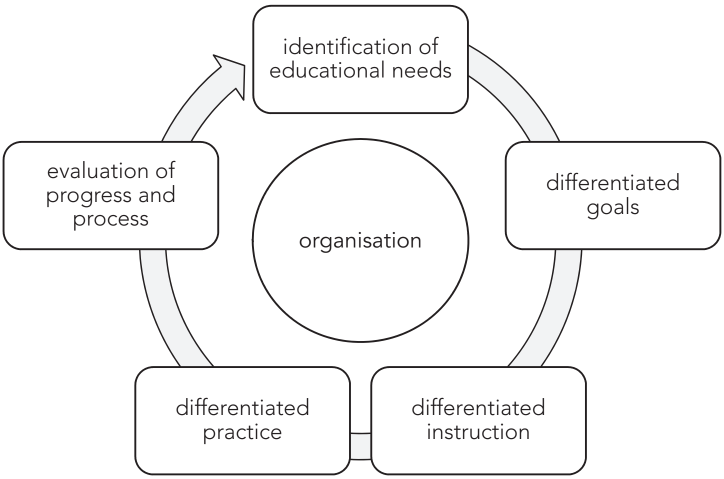


Figure 2.2 Cycle of differentiation.

Organisation is placed centrally in the cycle, because successful implementation of differentiation depends on a facilitative organisational structure and good classroom management. A key organisational characteristic of the model for differentiation agreed upon by the experts is the assignment of students to subgroups based on achievement level, allowing teachers to make instructional adaptations for subgroups of students with similar educational needs. Only the remaining individual educational needs that are not met within the subgroup call for individual accommodations.

The first step in the cycle of differentiation is the identification of educational needs. Initially, the teacher should assign students to subgroups (typically a low-achieving, an average-achieving and a high-achieving subgroup) based on their results on standardised tests and curriculum-based tests. In the course of the schoolyear, teachers continuously gather new and more detailed information about students' educational needs, for example with the analysis of daily work, informal observations and diagnostic conversations. The subgroups should be flexible, i.e. students should be able to switch groups based on changes in their educational needs.

Based on the educational needs of the students, differentiated goals should be set. Overarching objectives (the material that students should master at the end of primary school) and lesson goals (goals for a specific lesson) are distinguished. Overarching objectives should not only be formulated for average-achieving students but also specifically for low-achieving and high-achieving students (see also Appendix 2.1, differentiation in goals). The overarching objectives should be translated into concrete lesson goals, which are provided mostly by the curriculum. However, only some of the mathematics curricula available in the Netherlands differentiate lesson goals for three achievement levels. When the curriculum does not differentiate goals sufficiently, the teacher should formulate challenging but realistic lesson goals for all subgroups.

Based on the educational needs and the goals that have been set, the teacher differentiates instruction through broad whole-class instruction, subgroup instruction tailored to the needs of that subgroup, and individual adaptations. During whole-class instruction, the teacher should serve a broad range of educational needs by varying the difficulty level of questions, stimulating all children to think about the answer to a question by giving thinking time, teaching at various levels of abstraction, and using several input modalities (e.g. visual, verbal, tactile). In subgroup instruction, the teacher should adapt instruction to the educational needs of low-achieving and high-achieving students. It is assumed that, in general, low-achieving students need more guidance (e.g. explicit instruction) and instruction at lower levels of abstraction (e.g. using blocks to represent and calculate a sum) while high-achieving students need more exploratory instruction about advanced content with a focus on conceptual understanding (e.g. the relation between multiplication and division). To the extent possible, the teacher should also take into account individual differences during subgroup instruction and while giving individual feedback.

In the practice phase of the lesson, the subgroups need quantitatively and qualitatively different tasks. For the low-achieving subgroup, completing all regular tasks is often not realistic, so the tasks that are crucial for mastery of the objectives for low-achieving students should be selected (the remaining tasks can still be completed when students have time left). For the high-achieving subgroup, the regular material should be compacted. For the most popular Dutch mathematics curricula, guidelines exist to inform the teacher which tasks can be skipped by high-achieving students. In other cases, the teacher should remove most of the repetitive tasks and select the tasks that are crucial steps towards mastery of the objectives. The time freed up by compacting should be spent on enrichment. Some curricula provide enrichment tasks and these tasks can be used in some cases, but they are often not sufficiently challenging for very high-achieving students. Therefore, supplemental enrichment curricula should also be used. Technological applications such as mathematics websites and instructional computer programmes can also be valuable tools for individual differentiation, provided that they are used deliberately for additional practice in areas the student does not master yet or for enrichment at an appropriate challenge level.

The final step in the cycle of differentiation is the evaluation of progress and learning process. Based on daily work and achievement tests, the teacher should evaluate whether the students have met the lesson goals. Regarding process, the teacher should evaluate whether the applied adaptations of instruction and practice had the desired effect. For example, when a teacher has intentionally taught the low-achieving subgroup at a lower level of abstraction, the teacher should evaluate whether this was helpful for these students. To gauge the effectiveness of the accommodations made, the teacher may supplement achievement results with informal measures such as observations or diagnostic conversations. The evaluation phase informs the teacher about students' current achievement level and about instructional approaches that work for these students, completing the cycle and serving as new input for the identification of educational needs.

2.2.4 Concluding summary Study 1

The aim of Study 1 was to operationalise the concept of differentiation by achieving consensus among a consortium of mathematics experts about a coherent set of strategies for differentiating primary school mathematics education. A combination of focus group discussions and the Delphi technique was used to investigate the experiential knowledge of eleven experts in mathematics education systematically. Consensus was reached on all summarising concepts and on the majority of specific statements from the Delphi questionnaire. The input from the experts was synthesised into a cycle of differentiation consisting of the following five steps: identification of educational needs, differentiated goals, differentiated instruction, differentiated practice, and evaluation of learning progress and process. For each step, strategies were specified, providing teachers with concrete guidelines for implementing differentiation in primary school mathematics.

2.3 Study 2

2.3.1 Aims Study 2

Study 2 linked the results of Study 1 to teachers' daily practice by investigating teachers' self-assessed use of the strategies for differentiation recommended by the experts. Therefore, we developed the Differentiation Self-Assessment Questionnaire (DSAQ) which covers the recommended strategies in five subscales corresponding with the five steps of the cycle of differentiation (see also section 2.3.2.2). Since the DSAQ was newly developed, we also aimed to investigate its statistical properties, including its factor structure and relation with other scales.

The development of a new instrument was necessary to ensure coverage of the broad set of strategies recommended by the experts in Study 1. Another recently developed instrument to measure teachers' self-reported use of differentiation is the Differentiated Instruction Scale (DIS; Roy et al., 2013). This instrument was based on a similar theoretical framework and there is overlap between the content of the items of the DIS and the DSAQ. However, the DIS has only twelve items and is not sufficiently specific to measure all strategies recommended by the experts. Regarding progress monitoring, for example, the

DIS includes the rather general item 'analyse data about students' academic progress' while the DSAQ distinguishes between the different types of progress monitoring recommended by the experts, ranging from standardised tests to diagnostic interviews. Thus, the added value of the DSAQ is that it is a detailed measure of the specific strategies recommended by the experts in Study 1.

The first aim of the current study was to examine the factor structure of the DSAQ. The literature reviewed in the introduction indicates that effective differentiation entails two components: progress monitoring and instructional adaptations. The steps in the cycle of differentiation reflect these components: Identification of educational needs and evaluation of progress and process involve progress monitoring, while differentiated goals, instruction, and practice involve instructional adaptations. For the DIS, Roy et al. (2013) found that a model with these two factors provided a better fit to the data than a model in which all items loaded on one general differentiation factor. Therefore, we investigated whether the DSAQ has a similar factor structure by comparing the fit of a two-factor model with one factor for progress monitoring and one factor for instructional adaptations to the fit of a one-factor model.

The second aim was to examine the convergent and divergent validity of the DSAQ by investigating its relationship with other teacher self-report scales. Teacher self-efficacy is a multidimensional construct which comprises teachers' perceived ability to perform various aspects of teaching (Skaalvik & Skaalvik, 2007; Tschannen-Moran & Woolfolk Hoy, 2001). Theoretically, self-assessed usage of differentiation as measured by the DSAQ should be more closely related to aspects of self-efficacy related to differentiation than to other aspects of teacher self-efficacy. Specifically, we expected stronger correlations with scales that measure teachers' self-efficacy for instruction to students of diverse achievement levels, for adapting education to individual students' needs, and for self-assessed prerequisite knowledge for differentiation (which would support convergent validity) than with scales that measure teachers' self-efficacy for motivating students, for coping with changes and challenges, and for classroom management (which would support divergent validity).

The third and main aim was to investigate teachers' self-assessed use of the strategies for differentiation recommended by the experts. Besides examining teachers' overall usage, we also aimed to identify strategies which were relatively infrequently used. Such information may provide starting points for teacher professional development by indicating in which areas teachers perceive most room for improvement. Based on the literature reviewed in the theoretical background as well as on the input from the experts in Study 1, we hypothesised that average scores would be low to moderate and that specialised strategies aimed at individual students' unique educational needs as well as strategies targeted specifically at high-achieving students would be used relatively infrequently.

2.3.2 Method Study 2

2.3.2.1 Participants and procedure

The sample consisted of 268 primary school teachers working at 31 schools participating in a large-scale project about differentiation. Schools were informed about the project through flyers and advertisements and could register themselves for participation on a project website. The schools were located in rural and urban areas spread across the Netherlands and were diverse in terms of school size, student population, religious background, and mathematics curriculum used. All 325 teachers of Grade 1 through 6 of the participating schools were invited by email to fill out an online questionnaire containing the DSAQ and related scales. The questionnaire was administered at the beginning of the 2012 – 2013 school year. A total of 268 teachers (83%) completed the questionnaire and gave informed consent. The remaining teachers did not give informed consent (n = 3), completed the questionnaire only partly (n = 7) or did not respond at all (n = 47). On average, participants had 15.6 years of teaching experience (range 0 – 40 years). Seventy-one teachers (26%) taught a multigrade class. Fifty-four teachers (20%) worked full-time, whereas most teachers worked two, three or four days a week (61, 81 and 55 teachers respectively).

2.3.2.2 Instruments

The DSAQ was developed to examine how teachers assess their use of the strategies recommended by the experts in Study 1. Each subscale represents one step of the cycle of differentiation and covers core strategies for differentiation belonging to that step. Subscales and sample items are provided in Table 2.2. Organisational aspects of the model for differentiation were not captured in a separate scale but were partly covered in the subscales corresponding to each step of the cycle. In an earlier pilot study, a pilot version of the DSAQ had been completed by 27 teachers recruited at four schools. Based on the analysis of the internal consistency of the pilot version, which was acceptable to good, some adaptations were made in the final version of the DSAQ. The internal consistency of the final version, obtained in the current sample (N = 268), is reported in section 2.3.3.1.

To assess the convergent and divergent validity of the DSAQ, subscales from two well-established multidimensional teacher self-efficacy scales were selected. The Norwegian Teacher Self-Efficacy Scale (Skaalvik & Skaalvik, 2007) was developed with special attention for adapting education to individual educational needs. It consists of six subscales with acceptable to good internal consistency which load on six primary factors, which in turn load on a second-order factor for general teacher self-efficacy (Skaalvik & Skaalvik, 2007). For the current study, the subscales for Instruction – which emphasises instruction to students of diverse achievement levels – and Adapting Education to Individual Students'

Table 2.2 Sample items and descriptive statistics (N = 268) of the administered scales

Scale	Sample item	Response options	No. of items	α	*M*	*SD*
DSAQ						
Identification of educational needs[a]	I analyse the answers on curriculum-based tests to assess a student's educational needs	1 = does not apply to me at all, 5 = fully applies to me	5	.69	3.64	.55
Differentiated goals[a]	I set extra challenging goals for high-achieving students	1 – 5 as above	6	.79	3.78	.55
Differentiated instruction[a]	I adapt the level of abstraction of my instruction to the educational needs of the students	1 – 5 as above	7	.72	3.81	.42
Differentiated practice[a]	I select the most important elaboration activities for very low-achieving students	1 – 5 as above	8	.72	3.46	.55
Evaluation of progress and process[a]	I use diagnostic conversations to evaluate whether specific students have met the lesson goals	1 – 5 as above	7	.86	3.56	.57
Additional scales						
Instruction[b]	How certain are you that you can explain central themes in mathematics so that even the low-achieving students understand?	1 = not certain at all, 4 = absolutely certain	4	.74	3.13	.37
Adapting education to individual students' needs[b]	How certain are you that you can adapt instruction to the needs of low-achieving students while you also attend to the needs of other students in class?	1 - 4 as above	4	.78	2.91	.44
Coping with changes and challenges[b]	How certain are you that you can manage instruction regardless of how it is organised (working with subgroups, multigrade classes with 3 grades, etc.)?	1 - 4 as above	4	.76	2.99	.43
Motivating students[b]	How certain are you that you can get students to do their best even when working with difficult problems?	1 - 4 as above	4	.76	2.95	.42
Classroom management[c]	How much can you do to control disruptive behaviour in the classroom?	1 = nothing, 9 = very much	7	.92	7.17	.77
Prerequisite knowledge for differentiation[d]	I know the different solution strategies that are used by children	1 = does not apply to me at all, 5 = fully applies to me	10	.84	3.79	.41

[a] Newly developed DSAQ-scales. [b] Adapted from Skaalvik and Skaalvik (2007). [c] Taken from Goei, Bekebrede, and Bosma (2011). [d] Adapted from Nationaal Expertisecentrum Leerplanontwikkeling (2010).

Needs[1] were selected to assess the convergent validity of the DSAQ while the subscales for Motivating Students and for Coping with Changes and Challenges were selected to assess the divergent validity. The subscales were translated into Dutch and the wording was adapted to make the items domain-specific for mathematics instruction.

To further examine the divergent validity, the subscale for classroom management from the well-established Ohio State Teaching Efficacy Scale (OSTES; Tschannen-Moran & Woolfolk Hoy, 2001) was administered in Dutch translation (Goei, Bekebrede, & Bosma, 2011). The OSTES consists of three subscales which load on a first-order factor and on a general teaching efficacy factor. The subscales have demonstrated high internal consistency and can be used independently with in-service teachers (Tschannen-Moran & Woolfolk Hoy, 2001).

As a third potential support for convergent validity, a self-assessment scale about Prerequisite Knowledge for Differentiation was adapted from an informal scale that had already been used to assess the level of prior knowledge in professional development programmes (Nationaal Expertisecentrum Leerplanontwikkeling, 2010). Teachers self-assess the extent to which they already possess the knowledge necessary for implementing differentiated instruction.

2.3.2.3 Analyses

Because we wanted to compare the fit of two specific models based on theory and previous findings, we used confirmatory factor analysis (CFA) to investigate the factor structure of the DSAQ. We first tested a one-factor model in which all DSAQ subscales loaded on one general differentiation factor. Second, we tested a two-factor model with one factor for progress monitoring (subscales Identification of Educational Needs and Evaluation of Progress and Process) and one factor for instructional adaptations (subscales Differentiated Goals, Differentiated Instruction and Differentiated Practice). Version 7.3 of the Mplus statistical package (Muthén & Muthén, 1998-2012) was used. Model fit was evaluated with the chi-square statistic, the comparative fit index (*CFI*), the Tucker-Lewis Index (*TLI*), the root mean squared error of approximation (*RMSEA*), and the standardised root mean square residual (*SRMR*). Values above .95 for the *CFI* and *TLI* and values below .06 and .08 for the *RMSEA* and *SRMR*, respectively, indicate good model fit (Hu & Bentler, 1999). The maximum likelihood estimator was used. In the standardised solution, the variance of the factors was fixed to 1 so all factor loadings could be estimated freely.

Correlational analyses were performed to assess the convergent and divergent validity of the DSAQ. We expected moderate to strong positive correlations with

[1] Compared to the four-item NTSES subscale for Adapting Instruction to Individual Students' Needs, the added value of the DSAQ is that it provides a more detailed measure of self-assessed use of a range of differentiation strategies.

Prerequisite Knowledge for Differentiation, Adapting Education to Individual Students' Needs, and Instruction (for students of all achievement levels). For the latter two scales, we expected that correlations would be lower for the factor Progress Monitoring than for the factor Instructional Adaptations, because Progress Monitoring does not focus on the instructional phase. Regarding divergent validity, we hypothesised that DSAQ scores would be less strongly related to self-efficacy for Motivating Students, Coping with Changes and Challenges, and Classroom Management, although we still expected positive correlations because these dimensions of teacher self-efficacy can be helpful when implementing differentiation.

To identify areas of relatively low use, we compared the means of all single items to the mean of their factor. If a mean was more than one standard deviation below the mean of the factor, it was classified as relatively low.

2.3.3 Results Study 2

2.3.3.1 Properties of the DSAQ: Internal consistency and factor structure

As reported in Table 2.2, the internal consistencies of the DSAQ subscales were acceptable to good (Streiner, 2003).

The results of the confirmatory factor analysis indicate that the one-factor model in which all five subscales loaded on a general differentiation factor did not fit the data well: $\chi^2(5) = 55.126$, $p < .001$; *RMSEA* = .193 (90% CI .149 – .241); *CFI* = .912; *TLI* = .824; *SRMR* = .050. The two-factor model had a good fit: $\chi^2(4) = 5.637$, $p = .228$; *RMSEA* = .039 (90% CI .000 - .107); *CFI* = .997; *TLI* = .993; *SRMR* = .017. For the factor Progress Monitoring, standardised factor loadings were .84 (*SE* = 0.03, $R^2 = .71$) for Identification of Educational Needs and .85 (*SE* = 0.03, $R^2 = .73$) for Evaluation of Progress and Process. For the factor Instructional Adaptations, standardised factor loadings were .77 (*SE* = .04, $R^2 = .59$) for Differentiated Goals, .75 (*SE* = .04, $R^2 = .56$) for Differentiated Instruction, and .74 (*SE* = .04, $R^2 = .54$) for Differentiated Practice ($p < .001$ for all factor loadings). The correlation between the factors was .78 ($p < .001$). Since the two-factor model provided a better fit, the two factor scores (average of the subscale scores comprising that factor) were used in subsequent analyses.

2.3.3.2 Convergent and divergent validity: Correlations with other scales

The correlations between the two DSAQ factors and related scales are reported in Table 2.3. In support of convergent validity, the correlation with Prerequisite Knowledge for Differentiation was strong for both factors. As hypothesised, correlations with self-efficacy for Instruction and self-efficacy for Adapting Education to Individual Students' Needs were moderate to strong for the factor Instructional Adaptations and somewhat lower for Progress Monitoring. Regarding divergent validity, the correlations with Motivating

Table 2.3 Correlations ($p < .001$) between DSAQ factor scores and related scales

	Progress Monitoring		Instructional Adaptations	
Scale	*r*	95% CI	*r*	95% CI
Selected for convergent validity				
Prerequisite knowledge for differentiation	.62	.54 - .68	.70	.64 - .76
Instruction (to students of all achievement levels)	.40	.30 - .49	.47	.37 - .56
Adapting education to individual students' needs	.38	.28 - .48	.56	.48 - .64
Selected for divergent validity				
Motivating students	.30	.19 - .40	.37	.27 - .47
Classroom management	.34	.23 - .44	.40	.30 - .50
Coping with changes and challenges	.42	.31 - .52	.58	.49 - .65

Students and Classroom Management were less strong, although still in the moderate range. Contrary to expectations, Coping with Changes and Challenges correlated strongly with the factor Instructional Adaptations.

2.3.3.3 Distribution of DSAQ scores: Mean scores and infrequently reported strategies

The mean factor scores were 3.60 (*SD* = 0.52) for Progress Monitoring and 3.68 (*SD* = 0.43) for Instructional Adaptations. With a range from 1.83 to 4.86 for Progress Monitoring and from 2.45 to 4.94 for Instructional Adaptations, the factor scores were normally distributed at the high end of the scale. Table 2.2 provides the means and standard deviations of all subscales. Taken together, the mean factor and subscale scores reflect moderate to high self-assessed use of differentiation strategies.

Table 2.4 provides the means and standard deviations for each item of the DSAQ. Five items – numbers 3.7, 4.2, 4.4, 4.8, and 5.5 – had a mean score at least one standard deviation below the mean of their factor. Two of these - the use of diagnostic conversations to evaluate whether the learning goals have been met and the adaptation of type of practice to students' needs - reflect specialised strategies because they involve the refined diagnosis of and adaptation to individual students' needs. Other specialised strategies (items 1.5 and 5.7) also had somewhat lower means, although these means were within one standard deviation of the factor mean. The three remaining infrequently reported items concerned adaptations for high-achieving students, namely additional on-level instruction or guidance, curriculum compacting, and the use of computer programmes for additional challenge. Nevertheless, two other strategies targeted at high-achieving students (items 2.5 and 4.5) were frequently reported.

Table 2.4 Means and standard deviations of DSAQ items (scale range 1 - 5)

DSAQ item	*M*	*SD*
Subscale 1: Identification of educational needs		
1.1 I analyse the answers on curriculum-based tests to assess a student's educational needs	4.02	0.77
1.2 I analyse the answers on standardised tests to assess a student's educational needs	3.49	0.91
1.3 I assess specific students' educational needs based on daily maths work	3.75	0.72
1.4 I assess specific students' educational needs based on (informal) observations during the maths lesson	3.76	0.77
1.5 If necessary, I conduct diagnostic conversations to analyse the educational needs of specific students	3.20	0.90
Subscale 2: Differentiated goals		
2.1 I set different goals for the children, dependent on their achievement level	3.62	0.79
2.2 I set extra challenging goals for high-achieving students	3.57	0.83
2.3 I set well-considered minimum goals for very low-achieving students	3.75	0.76
2.4 I know the opportunities for differentiation offered by the curriculum	4.03	0.68
2.5 I use the opportunities the curriculum offers for differentiation for high-achieving students	3.88	0.84
2.6 I use the opportunities the curriculum offers for differentiation for low-achieving students	3.83	0.82
Subscale 3: Differentiated instruction		
3.1 I adapt the level of abstraction of instruction to the needs of the students	3.95	0.55
3.2 I adapt the modality of instruction (visual, verbal, manipulative) to the needs of the students	3.82	0.62
3.3 I adapt the pace of instruction to the needs of the students	3.95	0.56
3.4 I deliberately ask open-ended questions during whole-class instruction	3.82	0.67
3.5 I deliberately ask questions at various difficulty levels during whole-class instruction	3.69	0.73
3.6 I regularly provide low-achieving children with additional instruction (extended instruction, pre-teaching)	4.25	0.64
3.7 I regularly provide high-achieving students with additional instruction or guidance at their level, in a group or individually	3.20	0.92
Subscale 4: Differentiated practice		
4.1 I vary different types of practice during the maths lesson (e.g. individual or group work, solution spoken, written or drawn)	3.53	0.78
4.2 I adjust different types of practice to the needs of the students in the classroom (e.g. having a specific child complete exercises on the computer because this child learns more in this way)	3.04	0.83
4.3 I select the most important tasks for very low-achieving students	3.73	0.73
4.4 I use curriculum compacting for high-achieving students	3.20	1.25
4.5 I provide high-achieving students with enrichment tasks	4.00	0.87
4.6 I also use computer programmes or maths websites in my maths lessons	3.68	0.97
4.7 I use computer programmes and/or maths websites to offer children focused practice in a skill that they do not sufficiently master	3.32	0.96

Table 2.4 continues on next page

Table 2.4 *Continued*

DSAQ item	*M*	*SD*
4.8 I use computer programmes and/or maths websites to offer specific children additional challenge in the maths lesson	3.15	1.05
Subscale 5: Evaluation of progress and process		
5.1 I use scores on standardised and curriculum-based tests to evaluate whether the learning goals have been met	4.04	0.73
5.2 I analyse the answers on curriculum-based tests to evaluate whether the learning goals of that unit have been met	4.06	0.72
5.3 I regularly evaluate whether all students have met the learning goals based on their daily maths work	3.75	0.85
5.4 I evaluate whether all students have met the lesson goals based on (informal) observations during the maths lesson	3.45	0.86
5.5 I conduct diagnostic conversations to evaluate whether specific students have met the lesson goals	2.85	0.87
5.6 I evaluate whether the type of instruction and practice chosen by me were effective for the majority of the students in the class	3.44	0.77
5.7 I evaluate whether a specific type of instruction was effective for specific students	3.32	0.80

2.3.4 Concluding summary Study 2

Study 2 investigated teachers' self-assessed implementation of differentiation using the DSAQ. The first goal was to examine the psychometric properties of the DSAQ. The subscales of the DSAQ were internally consistent and loaded on two correlated but distinct factors: Progress Monitoring and Instructional Adaptations. Confirmatory factor analysis demonstrated that this two-factor structure provided a better fit than a one-factor model, which converges with the findings reported by Roy et al. (2013). The second goal was to examine the convergent and divergent validity of the DSAQ. The pattern of correlations between the DSAQ and other scales supported its convergent and divergent validity. As expected, strong to moderate correlations with Prerequisite Knowledge for Differentiation, Adapting Education to Individual Students' Needs, and Instruction were found. As hypothesised, the correlations with the scales selected for testing the divergent validity were lower, except for the correlation with Coping with Changes and Challenges which was unexpectedly strong.

The third and main goal was to examine teachers' perceived usage of the strategies recommended by the experts. With factor means in the moderate to high range, teachers assessed their use of differentiation strategies more highly than we had expected. Five items with relatively low means were identified. In support of our hypothesis, these items concerned specialised strategies and strategies targeted at high-achieving students.

2.4 General discussion

Teachers are required to implement differentiation for students of diverse achievement levels. However, the term differentiation had been used in diverse ways and the literature did not provide sufficient information regarding the most effective strategies to provide teachers with general guidelines for implementing differentiation. To fill this gap, Study 1 operationalised the concept of differentiation by achieving consensus among a consortium of experts about a model and strategies for differentiation in primary school mathematics. Study 2 investigated the degree to which Dutch teachers already implement the strategies suggested by the experts.

Study 1 resulted in a model for differentiation consisting of five steps: identification of educational needs, differentiated goals, differentiated instruction, differentiated practice, and evaluation of progress and process. These steps reflect the two core components of differentiated instruction identified by Roy et al. (2013). Progress monitoring is captured by the steps of identification of educational needs and evaluation of progress and process. The component of instructional adaptations is represented by the steps of differentiated goals, instruction, and practice. Study 2 demonstrated that a two-factor model in which the subscales of the DSAQ load on these two factors provides a better fit than a one-factor model. Our findings converge with the findings reported by Roy et al. (2013), supporting the idea that progress monitoring and instructional adaptations are two distinct but related components of differentiation.

New in this study is expert consensus on *how* progress should be monitored and how goals, instruction and practice should be adapted to the learning needs of students with diverse achievement levels. Regarding progress monitoring, the experts recommended to use standardised and curriculum-based tests first to divide students over achievement groups. More refined and informal measures such as the analysis of daily work should be used frequently to monitor short-term progress, to diagnose unique educational needs, and to determine whether a (temporary) adjustment of the groups is necessary. Compared to technological applications which tend to make use of one or two types of assessment to monitor progress (e.g. McDonald Connor et al., 2009; Ysseldyke & Tardrew, 2007), the experts recommended a broader range of strategies and indicated how they can be used together. The strategies have complementary purposes: While relatively formal and standardised tests are useful to get an overview of what a student can do, more informal and qualitative measures such as diagnostic conversations and the analysis of daily work provide valuable information about why a student struggles with a certain problem and what the student needs.

The use of within-class homogeneous achievement groups provides the opportunity to tailor subgroup instruction to similar educational needs and has demonstrated positive effects (Kulik & Kulik, 1992; Lou et al., 1996; Slavin, 1987; Tieso, 2005). In line with Slavin

(1987), the experts stressed the importance of flexibility, i.e. allowing students to switch between groups based on changes in their educational needs. The literature indicates that the effects of within-class ability grouping may depend upon student achievement level, with smaller or even negative effects for low-achieving students (Deunk et al., 2015; Lou et al., 1996). Nevertheless, the experts clearly perceived small-group instruction as a good way to provide low-achieving students with the instruction they specifically need. Also, students are only grouped for part of the lesson and participate in the whole-class instruction for students of all ability levels as well. Future research should establish whether these conditions ensure that low-achieving students also profit from this type of within-class flexible ability grouping.

Regarding instructional adaptations, the experts recommended a coherent set of strategies to differentiate goals, instruction and practice. This comprehensive approach is somewhat broader than technology-based interventions which have tended to focus on differentiation of either instruction (Individualizing Student Instruction) or practice (Accelerated Math). Many of the strategies recommended by the experts are supported by previous research, including the adaptation of practice tasks to the skill level of the student (Ysseldyke & Tardrew, 2007), the use of explicit instruction and visual representations for low-achieving students (Gersten et al., 2009) and the use of compacting, enrichment, and instruction at challenge level for advanced students (Rogers, 2007). To use teachers' time efficiently, the experts recommended to teach the whole class when possible, to use subgroups when the diverse educational needs of subgroups require this, and to serve remaining unique educational needs individually. Thus, the experts recommended both universal supports (supports for all students such as varying the difficulty level of questions in broad whole-class instruction) and targeted supports (supports specifically for low-achieving and high-achieving students including small-group instruction and differentiation in practice tasks). The experts also recommended some adaptations to individual students' educational needs (e.g. the adaptation of type of practice to the preference of specific students), but they realised that such specialised adaptations were advanced and primarily suitable for teachers who already master basic strategies for differentiation.

To link the advice provided by the experts to teachers' daily practice, Study 2 investigated teachers' self-reported usage of the recommended strategies. Overall, DSAQ scores were moderate to high, exceeding the expectations we had based on previous studies. Perhaps, the different context (primary schools in the Netherlands versus middle schools in the United States) can explain the discrepancy with the low use of differentiation strategies reported by Moon et al. (2002). Our findings are more similar to those of a recent study with Canadian primary school teachers in which moderate usage was reported (Roy et al., 2013). Nevertheless, the moderately high self-assessments in the current study seem discrepant with the finding of the Dutch Inspectorate of Education that adequate adaptations to students' diverse educational needs are only made at about half of the schools (Inspectie van het Onderwijs, 2012). Also, the experts in Study 1 clearly perceived a need for professional development about differentiation. Perhaps, the inspectors of education and the experts from our consortium have high standards for the *quality* of implementation which are not captured by the DSAQ. Teachers might also overestimate their implementation. Refined observational studies are necessary to examine whether teachers' high self-assessed usage of differentiation strategies can be confirmed by external observers.

In line with previous studies (McLeskey & Waldron, 2002, 2011; Reis et al., 2004; Scott et al., 1998; Westberg et al., 1993; Westberg & Daoust, 2003), specialised studies and strategies targeted at high-achieving students were used relatively infrequently. Two specialised strategies - the use of diagnostic conversations to evaluate whether the learning goals have been met and adaptation of the type of practice to specific students' needs - were relatively infrequently reported. This corresponds with the view expressed by the experts that individual-level differentiation is advanced and primarily suitable for teachers who already implement group-based strategies for differentiation successfully (Van Groenestijn, Borghouts, & Janssen, 2011).

Three strategies targeted at high-achieving students – curriculum compacting, the use of computer programmes for additional challenge, and targeted instruction for these students – were used infrequently. The difference between the use of instruction targeted at high-achieving students versus low-achieving students is especially striking. Perhaps, teachers are not aware that high-achieving students also need guidance when working on sufficiently challenging enrichment tasks (VanTassel-Baska & Stambaugh, 2005). Many teachers do implement some differentiation in practice tasks. However, there is still a lot of room for improvement, since it seems that only few teachers use a complete approach including challenging goals, curriculum compacting, enrichment tasks and on-level guidance. Low usage of differentiation for high-achieving students has repeatedly been attributed to a lack of the specific attitudes, knowledge, and skills this requires (Latz, Speirs Neumeister, Adams, & Pierce, 2009; Megay-Nespoli, 2001; VanTassel-Baska & Stambaugh, 2005). Many teacher educators feel that initial teacher training does not adequately prepare teachers to differentiate instruction for high-achieving students (Schram, Van der Meer, & Van Os, 2013). Thus, it seems important that this topic receives sufficient attention in teacher training and professional development programmes.

The following limitations should be taken into account. First, the results of consensus procedures are inherently restricted by the participating experts. The risk that other experts might have provided different input cannot be eliminated but was diminished in this study by recruiting experts working for several different institutions for both pre-service and in-service teacher training. Second, some experts missed some components of the procedure (a focus group discussion or a round of the Delphi questionnaire). This limitation was compensated for by the repetitive nature of the procedure: Participants who missed one component could still provide comments and additional input in the subsequent component. Third, it is possible that teachers provided socially desirable answers since a self-report questionnaire was used in Study 2. The variability between the items provides an indication that teachers did not simply rate themselves highly on all items to create a favourable impression. Nevertheless, we state again that observational studies are necessary to investigate how self-reported use is related to observed use of differentiation strategies. Fourth, it is unknown whether non-responders differed from teachers who did respond to the questionnaire, although the response rate of 83% is quite good.

A strong combination of methodologies was used. Study 1 employed an innovative methodology which combined the advantages of two methods: Focus group discussions are suitable for creating shared understanding and generating ideas, while the Delphi procedure gives all participants an equal and anonymous say in the systematic evaluation of those ideas. This combination was fruitful and efficient and we recommend it for future research. Moreover, the collaboration with experts who were familiar with the daily practice of teaching enhanced the feasibility of the findings. Based on their experience in various settings, the experts perceived these strategies as effective and feasible. In addition to this expert perspective, Study 2 examined the results of Study 1 from a teacher perspective. The fact that the teachers in Study 2 reported to use most of the strategies recommended by the experts in Study 1 shows that teachers acknowledge the need to differentiate and that the recommended strategies are largely compatible with teachers' daily practice. At the same time, the discrepancy between the teachers' and experts' perception of the degree of application of differentiation opens up an interesting avenue for future research. Thus, the inclusion of two complementary sources of information provides a richer perspective on differentiation. Despite these methodological advantages, future research is necessary to test empirically whether the implementation of the strategies recommended by the experts leads to higher student achievement.

The results of the current studies contribute to scientific research as well as to educational practice. Although the experts in Study 1 departed from a practical rather than a theoretical perspective, the elements of the cycle of differentiation overlap with elements of more general didactical models, such as Van Gelder's didactical analysis model (Van Gelder, Oudkerk Pool, Peters, & Sixma, 1973) and De Corte's didaxology (De Corte, Geerligs, Lagerweij, Peters, & Vandenberghe, 1976). Apparently, effective readiness-based differentiation is consistent with the principles of general good teaching, with the addition that each stage of teaching needs to be differentiated. A first theoretical implication is that, rather than studying specific elements (i.e. differentiated practice) in isolation, it seems promising to move towards an integral view of differentiation which involves all stages of teaching. Second, the experts clearly endorsed the view that students of different achievement levels have different educational needs and need different treatments at least part of the time, echoing the aptitude-treatment interaction literature (Cronbach & Snow, 1977). This emphasises the need to consider the potential variation between students in the design and analysis of educational intervention studies: What works for high-achieving students, may not work for low-achieving students and vice versa.

At the practical level, the model and strategies for differentiation recommended by the experts can be used in teacher training and professional development, for which purpose they have also been published in a Dutch journal for practitioners (Van de Weijer-Bergsma & Prast, 2013). Current educational policies require teachers to implement differentiation and our results provide teachers with concrete advice on how to do this. The cycle of differentiation can be used as a framework to structure professional development about differentiation. It shows teachers that differentiation requires attention at all stages of teaching in one coherent approach. The recommended strategies provide teachers with practical suggestions for each step (these can be found in section 2.2.3, Appendix 2.1, and also in the DSAQ-items listed in Table 2.4). The focus on mathematics promoted the concreteness of the results, since domain-specific guidelines can be applied directly without the need to transfer general principles. For example, the general guideline that advanced learners should be adequately challenged was made ready for use by providing achievement criteria for selecting high-performing students, suggestions for increasing task difficulty, guidelines for compacting, and a list of supplemental enrichment curricula. Nonetheless, the principles behind these concrete recommendations, including the cycle of differentiation, seem to be applicable in other domains as well. Future research could examine whether and how our results extend to other domains.

Study 2 provides researchers and practitioners with a new tool, the DSAQ. Researchers can use it, for example, as a pre- and post-assessment in intervention studies or to investigate teachers' self-assessed implementation in other countries. In professional development, the DSAQ can inform trainers about areas in which teachers perceive most room for improvement. Theoretically, Study 2 builds on the existing literature by providing support for the two-dimensional structure of differentiation. Moreover, this study is the first to investigate the self-reported use of a broad range of strategies for differentiation in mathematics in the Netherlands. The identified areas of low usage have practical implications, including the need to pay sufficient attention to differentiation for high-achieving students in teacher training and professional development.

References

Brown, J., & Morris, D. (2005). Meeting the needs of low spellers in a second-grade classroom. *Reading & Writing Quarterly*, *21*, 165-184. doi:10.1080/10573560590915969

Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, *43*, 161-173. doi:10.1080/00461520802178466

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: HarperPerennial.

De Corte, E., Geerligs, C., Lagerweij, N., Peters, J., & Vandenberghe, R. (1976). *Beknopte didaxologie* [Concise didaxology]. Groningen, the Netherlands: Wolters-Noordhoff.

Deunk, M., Doolaard, S., Smale-Jacobse, A., & Bosker, R. J. (2015). *Differentiation within and across classrooms: A systematic review of studies into the cognitive effects of differentiation practices*. Groningen, the Netherlands: GION.

Expertgroep doorlopende leerlijnen taal en rekenen (2008). *Over de drempels met taal en rekenen: Consolideren, onderhouden, uitbreiden en verdiepen* [Overcoming barriers in mathematics: Consolidating, maintaining, using and deepening]. Enschede, the Netherlands: Expertgroep doorlopende leerlijnen taal en rekenen. Retrieved from http://www.taalenrekenen.nl/referentiekader/rel_doc/downloads/Rekenrapport.pdf

Fuchs, L., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children*, *39*(5), 14-20. doi:10.1177/004005990703900503

Gelderblom, G. (2007). *Effectief omgaan met verschillen in het rekenonderwijs* [Dealing with differences in mathematics education effectively]. Amersfoort, the Netherlands: CPS.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, *79*, 1202-1242. doi:10.3102/0034654309334431

Goei, S. L., Bekebrede, J., & Bosma, T. (2011). *Teachers' Sense of Self Efficacy Scale: Meningen van leraren, experimentele versie* [Teachers' Sense of Self Efficacy Scale: Teacher's opinions, experimental version]. Amsterdam, the Netherlands: Onderwijscentrum Vrije Universiteit.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*, 1008-1015. doi:10.1046/j.1365-2648.2000.t01-1-01567.x

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55. doi:10.1080/10705519909540118

Inspectie van het Onderwijs. (2012). *De staat van het onderwijs: Onderwijsverslag 2010-2011* [The state of education: Educational report 2010-2011]. Utrecht, The Netherlands: Inspectie van het Onderwijs.

Janson, D., & Noteboom, A. (2004). *Compacten en verrijken van de rekenles voor (hoog)begaafde leerlingen in het basisonderwijs* [Compacting and enrichment of the mathematics lesson for highly able students in primary education]. Enschede, the Netherlands: SLO.

Johnsen, S. K., Haensly, P. A., Ryser, G. A. R., & Ford, R. F. (2002). Changing general education classroom practices to adapt for gifted students. *Gifted Child Quarterly*, *46*, 45-63. doi:10.1177/001698620204600105

Kulik, J. A., & Kulik, C. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly*, *36*, 73-77. doi:10.1177/001698629203600204

Landeta, J. (2006). Current validity of the delphi method in social sciences. *Technological Forecasting and Social Change*, *73*, 467-482. doi:10.1016/j.techfore.2005.09.002

Latz, A. O., Speirs Neumeister, K. L., Adams, C. M., & Pierce, R. L. (2009). Peer coaching to improve classroom differentiation: Perspectives from project CLUE. *Roeper Review*, *31*, 27-39. doi: 10.1080/02783190802527356

Liamputtong, P. (2011). *Focus group methodology: Principles and practice*. London, UK: Sage.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & D'Appolonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*, 423-458. doi: 10.3102/00346543066004423

McDonald Connor, C., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P. S., ... Schatschneider, C. (2011a). Testing the impact of child characteristics x instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly*, *46*, 189-221. doi:10.1598/RRQ.46.3.1

McDonald Connor, C., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science*, *315*, 464-465. doi:10.1126/science.1134513

McDonald Connor, C., Morrison, F. J., Schatschneider, C., Toste, J. R., Lundblom, E., Crowe, E. C., & Fishman, B. (2011b). Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness*, *4*, 173-207. doi:10.1080/19345747.2010.510179

McDonald Connor, C., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., ... Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Development*, *80*, 77-100. doi:10.1111/j.1467-8624.2008.01247.x

McLeskey, J., & Waldron, N. L. (2002). Inclusion and school change: Teacher perceptions regarding curricular and instructional adaptations. *Teacher Education and Special Education*, *25*, 41-54. doi:10.1177/088840640202500106

McLeskey, J., & Waldron, N. L. (2011). Educational programs for elementary students with learning disabilities: Can they be both effective and inclusive? *Learning Disabilities Research & Practice*, *26*, 48-57. doi:10.1111/j.1540-5826.2010.00324.x

Megay-Nespoli, K. (2001). Beliefs and attitudes of novice teachers regarding instruction of academically talented learners. *Roeper Review*, *23*, 178-182. doi:10.1080/02783190109554092

Moon, T. R., Callahan, C. M., Tomlinson, C. A., & Miller, E. M. (2002). *Middle school classrooms: Teachers' reported practices and student perceptions* (No. RM02164). Storrs, CT: The National Research Center on the Gifted and Talented.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nationaal Expertisecentrum Leerplanontwikkeling (2010). *Professionele ontwikkeling: Tussenmeting project 'Als je merkt dat het werkt'* [Professional development: Midtime measurement project 'When you notice it works']. Enschede, the Netherlands: SLO.

Nurmi, J., Kiuru, N., Lerkkanen, M., Niemi, P., Poikkeus, A., Ahonen, T., ... Lyyra, A. (2013). Teachers adapt their instruction in reading according to individual children's literacy skills. *Learning and Individual Differences*, *23*, 72-79. doi:http://dx.doi.org/10.1016/j.lindif.2012.07.012

Reis, S. M., Gubbins, E. J., Richards, S., Briggs, C. J., Jacobs, J. K., Eckert, R. D., ... Schreiber, F. J. (2004). Reading instruction for talented readers: Case studies documenting few opportunities for continuous progress. *Gifted Child Quarterly*, *48*, 315-338. doi:10.1177/001698620404800406

Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M., & Kaniskan, R. B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal*, *48*, 462-501. doi:10.3102/0002831210382891

Rogers, K. B. (2007). Lessons learned about educating the gifted and talented: A synthesis of the research on educational practice. *Gifted Child Quarterly*, *51*, 382-396. doi:10.1177/0016986207306324

Roy, A., Guay, F., & Valois, P. (2013). Teaching to address diverse learning needs: Development and validation of a differentiated instruction scale. *International Journal of Inclusive Education*, *17*, 1186-1204. doi:10.1080/13603116.2012.743604

Schram, E., Van der Meer, F., & Van Os, S. (2013). *Omgaan met verschillen: (G)een kwestie van maatwerk* [Responding to differences: (Not) a matter of customisation] (No. AN2.6547.542). Enschede, the Netherlands: SLO.

Scott, B. J., Vitale, M. R., & Masten, W. G. (1998). Implementing instructional adaptations for students with disabilities in inclusive classrooms: A literature review. *Remedial and Special Education*, *19*, 106-119. doi:10.1177/074193259801900205

Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, *99*, 611-625. doi:10.1037/0022-0663.99.3.611

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. *Review of Educational Research*, *57*, 293-336. doi:10.3102/00346543057003293

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*, 99-103. doi:10.1207/S15327752JPA8001_18

Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D., & Yu, F. (2006). The international system for teacher observation and feedback: Evolution of an international study of teacher effectiveness constructs. *Educational Research & Evaluation*, *12*, 561-582. doi:10.1080/13803610600874067

Tieso, C. L. (2003). Ability grouping is not just tracking anymore. *Roeper Review*, *26*, 29-36. doi:10.1080/02783190309554236

Tieso, C. L. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted*, *29*, 60-89.

Tomlinson, C. A. (2005). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). Upper Saddle River, NJ: Pearson Education.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., ... Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, *27*, 119-145. doi:10.1177/016235320302700203

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*, 783-805. doi:10.1016/S0742-051X(01)00036-1

Van de Weijer-Bergsma, E., & Prast, E. J. (2013). Gedifferentieerd primair rekenonderwijs volgens experts: De resultaten uit een Delphi-onderzoek [Differentiated primary math education according to experts: Results from a Delphi study]. *Orthopedagogiek: Onderzoek en Praktijk*, *52*, 336-349.

Van Gelder, L., Oudkerk Pool, T., Peters, J., & Sixma, J. (1973). *Didactische analyse* [Didactical analysis]. Groningen, The Netherlands: Wolters-Noordhoff.

Van Groenestijn, M., Borghouts, C., & Janssen, C. (2011). *Protocol ernstige rekenwiskunde-problemen en dyscalculie* [Protocol severe mathematics difficulties and dyscalculia]. Assen, the Netherlands: Van Gorcum.

VanTassel-Baska, J., Feng, A. X., Brown, E., Bracken, B., Stambaugh, T., French, H., ... Bai, W. (2008). A study of differentiated instructional change over 3 years. *Gifted Child Quarterly*, *52*, 297-312. doi:10.1177/0016986208321809

VanTassel-Baska, J., & Stambaugh, T. (2005). Challenges and possibilities for serving gifted learners in the regular classroom. *Theory into Practice*, *44*, 211-217. doi:10.1207/s15430421tip4403_5

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Westberg, K. L., Archambault, F. X., Dobyns, S. M., & Salvin, T. J. (1993). The classroom practices observation study. *Journal for the Education of the Gifted*, *16*, 120-146. doi:10.1177/016235329301600204

Westberg, K. L., & Daoust, M. E. (2003). *The results of the replication of the classroom practices survey replication in two states*. Retrieved from http://www.gifted.uconn.edu/nrcgt/newsletter/fall03/fall03.pdf

Ysseldyke, J., & Bolt, D. M. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. *School Psychology Review*, *36*, 453-467. Retrieved from http://www.nasponline.org/publications/spr/abstract.aspx?ID=1847

Ysseldyke, J., Spicuzza, R., Kosciolek, S., & Boys, C. (2003). Effects of a learning information system on mathematics achievement and classroom structure. *The Journal of Educational Research*, *96*, 163-173. doi:10.1080/00220670309598804

Ysseldyke, J., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk*, *8*, 247-265. doi: 10.1207/S15327671ESPR0802_4

Ysseldyke, J., & Tardrew, S. (2007). Use of a progress monitoring system to enable teachers to differentiate mathematics instruction. *Journal of Applied School Psychology*, *24*, 1-28. doi: 10.1300/J370v24n01_01

Appendix 2.1
Summarising concepts included in the Delphi questionnaire

What follows are the translations of the concepts as they were included in the Delphi questionnaire. Background information that might be relevant for non-Dutch readers is given in the footnotes.

Organisation

The starting point is convergent differentiation2. Students are assigned to one of three subgroups based on standardised tests and / or curriculum-based tests. If curriculum-based tests are used to assess what students already master, the test score of the previous unit can be used, but an alternative is to use the end-of-unit test of the upcoming unit as a pretest. The teacher can change the grouping arrangement for a certain unit or lesson based on test scores. During mathematics classes, whole-class instruction, instruction to one of the subgroups (of low-achieving or high-achieving students) and independent practice are alternated. Average achievers take part in the whole-class instruction and receive individual feedback or guidance during the time for independent practice.

Differentiation in instruction

During whole-class instruction, the teacher serves different levels and educational needs to the extent that this is possible. The teacher can do this by teaching at different levels of abstraction and showing the connection between these different levels. The teacher should ask questions at varying difficulty levels, implying that some questions may be too easy or too difficult for some of the students in the class. During instruction to a subgroup, the teacher takes into account the educational needs of the students in that specific subgroup. For example, the teacher spends more time on lower levels of abstraction when teaching the low-achieving subgroup, while instruction to the high-achieving subgroup is mainly at a high level of abstraction. Additionally, it is assumed that the low-achieving subgroup needs more guidance (more direct instruction) than the high-achieving subgroup (more exploratory instruction). To the extent possible, the teacher also takes into account individual differences within a subgroup. For example, the teacher can accommodate to a student's need to verbalise a solution strategy himself, or a student's need for visualisation. An additional strategy for differentiation in instruction is the use of instructional videos.

[2] In the Netherlands, a distinction is often made between convergent and divergent differentiation (Gelderblom, 2007). In convergent differentiation, all student in a classroom work on roughly the same topics at the same time (even if they might engage in the topic at varying levels of complexity). In divergent differentiation, different students work on different learning goals and topics at the same time.

Differentiation in goals

A strong awareness of the learning trajectories and accompanying educational goals is essential for a good lesson. For differentiated education, this means that different goals are set for different students. Goals are differentiated primarily at the subgroup level. Highly competent teachers can also differentiate goals on an individual basis based on their professional insight. For the low-achieving subgroup, the objective is to master the fundamental level (1F)[3] at the end of primary school. For the average-achieving subgroup, the objective is to master the target level (1S) at the end of primary school. For the high-achieving subgroup, mastery of the target level is a minimum requirement, but additionally, more advanced goals (for example regarding logical reasoning) are set for these students. The goals for the end of primary school are converted into specific lesson goals for the three subgroups based on the curriculum and the professional insight of the teacher. These lesson goals should be both ambitious and realistic. The teacher keeps in mind the lesson goals while preparing and teaching his lesson. After the lesson, the teacher evaluates whether lesson goals have been met.

Differentiation in the practice phase

The different subgroups need quantitatively and / or qualitatively different practice tasks. From the tasks that the curriculum offers, the tasks at the minimum and fundamental level are most important for low-achieving students. The high-achieving subgroup can skip a large proportion of the tasks at minimum and fundamental level. Existing guidelines for compacting[4] can be used to select the tasks that high-achieving students do need to do. High achieving students spend the time that is freed up by compacting the regular material on enrichment. The enrichment tasks provided in the regular curriculum are often not sufficiently challenging, especially for gifted students. Additional enrichment should be provided for these students. Such enrichment may include assignments for which students have to carry out research or use information from different sources. Besides the adaptation (selection and supplementation) of tasks, practice can also be differentiated during instruction to subgroups. For example, the teacher could use the extended instruction for low-achieving students to solve the exercises together step-by-step, while a discussion of the big ideas behind a certain task may be more useful in the high-achieving subgroup.

---

3 In the Dutch educational system, overarching objectives (comparable to the common core state standards employed in the US) have been defined at two levels: The fundamental level (1F) that should be reached by all students and the target level (1S) that should be reached by about 65% of the students (Expertgroep doorlopende leerlijnen taal en rekenen, 2008).

4 An educational advisory company has published guidelines for compacting the most popular Dutch mathematics curricula (Janson & Noteboom, 2004). Children for whom the material should be compacted receive an additional booklet with an overview per lesson of the exercises they should do and the exercises they can skip.

Differentiation based on learning styles

The educational needs of different students may also vary within subgroups. For example, students may have a preference for certain formats (whole-class instruction, working together, working individually) and certain input modes (visual or verbal, written or spoken). It has been mentioned repeatedly that it is important for some students to express the content themselves or to have the content explained to them by another student. Teachers need to be aware of these differences between students and learn how they can vary their instruction, tasks and formats to accommodate various educational needs. Especially during the instruction to subgroups the teacher can accommodate to individual educational needs, provided that he is able to identify what kind of instruction or type of task a student needs to understand the content.

Differentiation in kindergarten

When the files of students with problematically low mathematics achievement in primary school are examined, it often turns out that problems with preparatory mathematics were already detected in kindergarten but that no or insufficient action has been taken to tackle those problems in the meantime. Factors that may play a role in this lack of action are beliefs of the teacher ('the child is not ready for preparatory mathematics' or 'children of this age should be allowed to play'), inadequate communication to the teacher of the next grade, and lack of knowledge of ways to tackle low achievement in (preparatory) mathematics. In order to respond more quickly to early signals of problems with acquiring preparatory mathematics skills, teachers should (a) set more specific and ambitious goals (what should the child be able to do at the beginning of grade 1?), (b) be more knowledgeable about levels of abstraction and be able to demonstrate the connections between various levels of abstraction, (c) realise that learning can take place in the process of playing if the activity is well adapted to the child's educational needs, and (d) that certain children need some additional instruction, also in kindergarten. Additionally, more attention should be given to providing extra challenge to students with highly developed preparatory mathematics skills.


53 - 2
-3
33

3

Teaching students with diverse achievement levels: Observed implementation of differentiation in primary mathematics education

Emilie J. Prast
Eva Van de Weijer-Bergsma
Evelyn H. Kroesbergen
Johannes E. H. Van Luit

Manuscript submitted for publication

Author contributions: All authors designed research; EvdW and EP collected the data under supervision of HvL and EK; EP analysed the data and wrote the paper; EvdW, EK and HvL provided feedback on the paper.

Abstract

Teachers have an important role in adapting education to the diverse educational needs of their students ('differentiation'). A detailed observation instrument was used to examine teachers' (N = 74) use of strategies for achievement-based differentiation in primary mathematics education, both before and after professional development (PD). At baseline, most teachers implemented some aspects of differentiation, but refined adaptations to the specific needs of low-achieving and high-achieving students were observed relatively infrequently. Teachers participated in a yearlong PD programme about differentiation in primary mathematics education. The PD had no observable short-term effects. Experimental teachers did implement more differentiation than control teachers in the year after the PD, which might be interpreted as a long-term effect, but alternative explanations could not be ruled out.

3.1 Introduction

Different learners have different needs. Teachers have an important role in adapting education to the diverse educational needs of their students (differentiation; Tomlinson et al., 2003). However, implementing differentiation is complex and requires specific knowledge and skills (adaptive teaching competence; Vogt & Rogalla, 2009). There are concerns that teachers do not sufficiently possess these skills and therefore do not implement differentiation optimally (e.g., Hertberg-Davis, 2009; Inspectorate of Education, 2012; Schumm, Moody, & Vaughn, 2000). This study is part of the large-scale project GROW (in Dutch, an acronym for differentiated mathematics education; see also Prast, Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2015), in which a professional development (PD) programme about differentiation in primary mathematics education is developed and evaluated. In the current study, we examine teachers' use of strategies for differentiation for students of diverse achievement levels in mathematics using a detailed observation instrument. Second, we investigate the effect of a PD programme on teachers' observed implementation of these strategies.

3.1.1 Readiness-based differentiation

Tomlinson and colleagues (2003, p.120) have defined differentiation broadly as 'an approach to teaching in which teachers proactively modify curricula, teaching methods, resources, learning activities, and student products to address the diverse needs of individual students and small groups of students to maximize the learning opportunity for each student in a classroom'. A distinction is made between differentiation based on student readiness (readiness for learning based on the current level of knowledge and skills), interest (preferred topics), and learning profile (preferred ways of learning) (Tomlinson et al., 2003). In the current study, we focus on differentiation based on student readiness, also called achievement-based differentiation, cognitive differentiation, or ability-based differentiation. Readiness-based differentiation is especially important in primary schools, because students within primary school classrooms vary widely in terms of their academic ability and achievement level (Deunk, Doolaard, Smale-Jacobse, & Bosker, 2015). Therefore, high-achieving and low-achieving students are likely to have different zones of proximal development (Vygotsky, 1978). To allow all students to work in their zone of proximal development and to be appropriately challenged, students need different instructional treatments (Cronbach & Snow, 1977). Therefore, teachers should adapt education to the diverse educational needs of their students (Corno, 2008). Readiness-based differentiation is aimed at students of *all* achievement levels (Roy, Guay, & Valois, 2013).

Roy et al. (2013) define readiness-based differentiation in terms of two components: academic progress monitoring and instructional adaptations. Thus, teachers should monitor students' academic progress and use this information to adapt instruction to students' current educational needs. Differentiated instruction is often (but not always) organised using within-class homogeneous achievement groups and may involve adaptations such as additional small-group instruction for low-achieving students as well as curriculum compacting and enrichment for high-achieving students. However, the way in which progress is monitored and the nature of instructional adaptations vary widely across studies (Prast et al., 2015).

Given the diverse uses of the term 'differentiation' in the literature, we developed a prescriptive model for differentiation based on expert consensus (Prast et al., 2015). We focused exclusively on the domain of primary mathematics education, since differentiation requires domain-specific pedagogical content knowledge (Shulman, 1986; Vogt & Rogalla, 2009). This resulted in the cycle of differentiation displayed in Figure 3.1 (Prast et al., 2015).

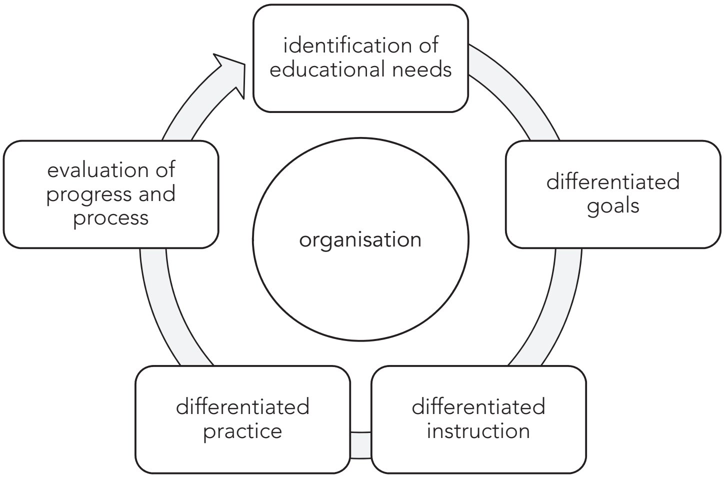


Figure 3.1 Cycle of differentiation (Prast et al., 2015; reprinted with permission).

The first step in the cycle of differentiation is the identification of educational needs. Based on formal and informal assessments, the teacher should assign students to three (or more) flexible homogeneous within-class achievement groups[1] based on students' current achievement level (Tieso, 2003, 2005). The achievement groups should be used part of the time to cater specifically to the educational needs of the different subgroups. Frequent

[1] We use 'achievement grouping' rather than 'ability grouping' since the groups are formed based on students' current academic achievement level which is dynamic, rather than reflecting a fixed academic ability level.

(informal) assessments should be used to signal when students should switch between achievement groups, and to gather more refined information about students' educational needs. In the second step, the teacher should set differentiated goals for the different subgroups. These goals should be challenging but realistic given the current achievement level of the students (Csikszentmihalyi, 1990). Third, based on the educational needs and the goals that have been set, the teacher should differentiate instruction through broad whole-class instruction which engages students of diverse achievement levels, subgroup instruction tailored to the needs of that subgroup, and individual adaptations. Besides additional instruction for low-achieving students, specific attention for high-achieving students is necessary (VanTassel-Baska & Stambaugh, 2005). Fourth, the practice tasks should be differentiated. For the low-achieving subgroup, the most crucial tasks towards mastery of the goals should be selected. For the high-achieving subgroup, the regular material should be compacted and supplemented with challenging enrichment tasks (Rogers, 2007; VanTassel-Baska & Wood, 2010). Finally, the teacher should use a range of formal and informal assessments to evaluate whether the students have met the goals and whether the applied instructional adaptations of instruction and practice had the desired effect. The evaluation phase informs the teacher about students' current achievement level and about instructional approaches that work for these students, completing the cycle and serving as new input for the identification of educational needs.

In the domain of mathematics, two important models for differentiating the instruction and practice phase are the 'stages of learning mathematics model' and the 'levels of abstraction model' (Van Groenestijn, Borghouts, & Janssen, 2011). The stages of learning mathematics model assumes that learning mathematics starts by building understanding of mathematical concepts and procedures (e.g., addition). After learning how to solve a particular type of task (e.g., how to add up two numbers) students develop fluency in solving those tasks (e.g., quick calculation or memorisation of basic addition sums). In the highest stage of mastery, students can apply their mathematical knowledge and skills flexibly (e.g., by selecting and combining appropriate strategies to solve a real-life problem). The development of these stages is not always linear and teachers should use this model flexibly, for example by moving back to the stage of building understanding when a student struggles to develop solution procedures. One way of building understanding is by teaching at a lower level of abstraction. The levels of abstraction model specifies four levels of abstraction: real-life action (e.g., counting blocks), concrete representation (e.g., pictures of blocks), abstract representation (de-contextualised schematic representations, e.g., tallies), and formal reasoning using abstract mathematical language (e.g., 5 + 2 =). According to the model, understanding a topic at a low level of abstraction helps students to understand tasks about the same topic at higher levels of abstraction. Teachers can use these models for differentiation in three ways: (1) to broaden whole-class instruction by teaching at multiple levels and showing the connection between the levels, (2) to help low-achieving students by diagnosing at which level they understand the task and moving back to a more fundamental level as necessary, and (3) to challenge high-achieving students by providing them with questions or tasks at a higher level.

3.1.2 Implementation of differentiation

The cycle of differentiation and the associated models and strategies illustrate that implementing differentiation requires advanced knowledge and skills. In line with Vogt and Rogalla (2009), we use the term 'adaptive teaching competency' to refer to teachers' capacities for making adaptations to students' identified educational needs. Adaptive teaching competency includes four dimensions: subject matter knowledge, the ability to diagnose students' current understanding and achievement, the ability to use diverse teaching methods to meet diverse students' needs, and classroom management skills (Vogt & Rogalla, 2009). Thus, it requires both general pedagogical skills (e.g., classroom management) and domain-specific subject matter knowledge (e.g., in mathematics) as well as pedagogical content knowledge (e.g., didactical models). Underlying these knowledge and skills, a positive attitude towards differentiation is required (Smeets, Ledoux, Regtvoort, Felix, & Mol Lous, 2015). There are widespread concerns that many teachers do not possess these attitudes, knowledge and skills sufficiently and therefore are insufficiently able to implement differentiation successfully (Hertberg-Davis, 2009; Inspectorate of Education, 2012; Schumm et al., 2000).

Previous findings regarding the implementation of differentiation are somewhat mixed. Early studies typically focused on one end of the achievement spectrum. These studies generally showed limited use of instructional adaptations for students with a learning disability (LD; reviewed by Scott, Vitale, & Masten, 1998) and for advanced students (Westberg & Daoust, 2003; Westberg, Archambault, Dobyns, & Salvin, 1993), across observational and self-report designs. Three recent studies examined teachers' self-reported usage of several strategies for differentiation for the full range of achievement levels in primary school (Prast et al., 2015; Roy et al., 2013; Wan & Wan, 2013). In contrast to previous findings, teachers reported moderate to high implementation of differentiation in all of these studies (as indicated by total mean scores above the midpoint of the scale), although some strategies were reported less frequently than others. Observation instruments may be a more objective indicator of actual use of differentiation strategies than self-report questionnaires. However, we did not find previous observational studies regarding the implementation of differentiation for students of all achievement levels in primary school. Some observational studies did examine differentiation in other settings (secondary school; Maeng & Bell, 2015) or for particular groups of students (students with LD: McKenna, Shin, & Shiullo, 2015; advanced students: Brighton, Moon & Huang, 2015) and these studies generally indicated limited application of differentiation. Differences between self-report and observational studies may indicate that teachers overestimate their own implementation of differentiation. Alternatively, it is possible that general differentiation strategies for a broad range of achievement levels are applied more frequently than strategies targeted specifically at the unique needs of students with LD or giftedness, or that differentiation is more commonly used in primary than in secondary schools. More observational research is necessary to clarify this issue.

3.1.3 Effects of PD on the implementation of differentiation

To the best of our knowledge, no previous studies have examined the effects of PD about differentiation for students of *all* achievement levels on teachers' implementation of differentiation using a standardised observation instrument. However, some studies have investigated this using other measures. Two self-report studies found that even brief PD (up to two days) had positive effects on teachers' self-efficacy for and self-reported use of differentiation, but these self-report studies were methodologically limited by a very basic measure of applied differentiation ("How often do you differentiate in your classroom?"; Dixon et al., 2014) and lack of a control group (Edwards, Carr, & Siegel, 2006). In contrast, a study using more objective measures (standardised tasks based on a video and vignette example; Vogt & Rogalla, 2009) found only partially positive effects after an extensive coaching trajectory (positive effects on lesson planning but not on instructional adaptations in direct response to students' needs).

Studies using standardised observation instruments have focused on adaptations for either low-achieving *or* high-achieving students within heterogeneous classrooms. Two studies about appropriately challenging instruction for advanced students (Johnsen, Haensly, Ryser, & Ford, 2002; VanTassel-Baska et al., 2008) found positive effects on teachers' implementation of differentiation, but two to three years of intensive PD were necessary to achieve this. A study about differentiation for students with LD found that, besides intensive PD, administrative support was important for successful implementation (Klingner, Ahwee, Pilonieta, & Menendez, 2003).

To sum up, while self-report studies have indicated that brief interventions can already enhance teachers' self-efficacy for implementing differentiation, observational studies have shown that changing teachers' observable behaviour is challenging. Studies about adaptations for students with LD or giftedness have shown that increasing the implementation of differentiation for these specific groups of students is possible with long-term training and extensive supports. However, there is still a need for observational studies about differentiation for students of the full range of achievement levels.

In the current study, we first explore teachers' baseline level of implementation of a range of strategies for readiness-based differentiation in mathematics using a detailed observation instrument. Second, we examine the effects of a teacher PD programme about differentiation in mathematics using the cycle of differentiation on teachers' observed implementation of these strategies. We hypothesise that the PD programme will have positive effects on teachers' implementation of differentiation.

3.2 Method

3.2.1 Design

The design of the study is shown in Table 3.1. Participating schools were randomly assigned to one of three cohorts: Cohort 1 participated in the PD programme in Year 1, Cohort 2 participated in the PD programme in Year 2, and Cohort 3 served as a control condition in both years (but was offered to participate in the PD programme in the following schoolyear).

Table 3.1 Research design

	Year 1 (2012–2013)		Year 2 (2013–2014)	
Cohort 1	PD programme		Follow-up	
Cohort 2	Control		PD programme	
Cohort 3	Control		Control	
Video observations	pre	post	pre	post

3.2.2 Participants

Data were collected in the context of the large-scale project GROW (see also Prast et al., 2015). 32 primary schools voluntarily signed up for participation, which involved free participation in the PD programme and two years of data collection. In the course of the project, two of these schools dropped out. The first school (assigned to Cohort 1), dropped out after the first measurement occasion because it perceived the project as too intensive and was excluded from the analyses. The second school (assigned to Cohort 2), quit with the PD programme in the course of Year 2 after identifying other priorities for PD. Only the Year 1 data collected at this school were included in the analyses. The participating schools were geographically spread across the Netherlands and were diverse in terms of school size (*M* = 208 students per school, range 52 to 550) and mathematics curriculum used (five different curricula in different versions).

For the current study, a subsample of teachers was selected for video observations to evaluate the effect of the PD programme on teachers' observed application of differentiation. At each school, the teacher of grade 3 was observed (in case of multigrade classes, the class including grade 3 (e.g., grade 2–3) was selected; in individual cases in which observing the grade 3 teacher was not feasible (e.g., because grade 3 was taught by a temporary replacement teacher), another grade was observed). Grade 3 was chosen because it was in the middle of the range of grades in the sample (1 through 6). Additionally, three schools of Cohort 1 and three schools of Cohort 2 were observed more intensively, with planned video observations for all teachers of all grades. Thus, video observations were planned for 73 teachers in Year 1 and 70 teachers in Year 2. For some of these teachers, video observations were only planned in Year 1 (n = 17 teachers) or in Year 2 (n = 15 teachers), mainly due to teachers switching between grades (e.g., teaching grade 3 in one year only). Per timepoint and teacher, two video observations were planned. After data collection, some teachers for whom video observations had been planned had to be excluded because they had missed more than two PD meetings (one teacher in Year 1 and three teachers in Year 2) or because they did not provide data at pretest or posttest (18 teachers in Year 1 and 10 teachers in Year 2). Frequent reasons for missingness at one timepoint were long-term absence due to illness or maternity leave and technical problems with the videos. Thus, the final sample consisted of 55 teachers in Year 1 (75.3% of teachers for whom observations were planned) and 59 teachers (84.3%) in Year 2, for a total of 74 unique teachers.

Characteristics of the sample are described in Table 3.2. At the beginning of the study, teachers had an average of approximately 14 years of teaching experience (range

Table 3.2 Teacher characteristics

	Cohort 1 *M* (*SD*)	Cohort 2 *M* (*SD*)	Cohort 3 *M* (*SD*)	Total *M* (*SD*)
Years of experience	15.44 (12.27)	11.69 (9.15)	15.33 (13.21)	13.87 (11.24)
Grade level taught in Year 1	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Grade 1	3 (13.0%)	3 (13.6%)	0 (0.0%)	6 (10.9%)
Grade 2	3 (13.0%)	3 (13.6%)	1 (10.0%)	7 (12.7%)
Grade 3	6 (26.1%)	3 (13.6%)	5 (50.0%)	14 (25.5%)
Grade 4	3 (13.0%)	3 (13.6%)	0 (0.0%)	6 (10.9%)
Grade 5	2 (8.7%)	3 (13.6%)	1 (10.0%)	6 (10.9%)
Grade 6	3 (13.0%)	3 (13.6%)	1 (10.0%)	7 (12.7%)
Multigrade class	3 (13.0%)	4 (18.2%)	2 (20.0%)	9 (16.4%)
Grade level taught in Year 2	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Grade 1	3 (12.5%)	5 (20.0%)	0 (0.0%)	8 (13.6%)
Grade 2	2 (8.3%)	3 (12.0%)	0 (0.0%)	5 (8.5%)
Grade 3	3 (12.5%)	4 (16.0%)	6 (60.0%)	13 (22.0%)
Grade 4	3 (12.5%)	3 (12.0%)	0 (0.0%)	6 (10.2%)
Grade 5	2 (8.3%)	3 (12.0%)	0 (0.0%)	5 (8.5%)
Grade 6	2 (8.3%)	3 (12.0%)	0 (0.0%)	5 (8.5%)
Multigrade class	9 (37.5%)	4 (16.0%)	4 (40.0%)	17 (28.8%)

1 to 41 years). All grade levels were represented, with relatively more teachers teaching grade 3 in accordance with our sampling procedure. In Year 1 and Year 2 respectively, 16.4 and 28.8% of teachers taught a multigrade class (typically two, sometimes three adjacent grade levels). Differences in background variables between cohorts or years were no cause for concern, since the observed level of differentiation (DMI total score at pretest Year 1) was not significantly related to grade level (r = -.03, p = .822) nor did it differ between teachers of single versus multigrade classes (F (1, 53) = 0.42, p = 0.519). Although the correlation with years of experience approached significance (r = 0.24, p = .085), the scatterplot showed no simple linear effect (possibly curvilinear) and years of experience did not differ between cohorts (F (2, 67) = 0.93, p = 0.399).

3.2.3 Professional development programme

Ten three-hour team meetings spread across the schoolyear were provided for all teachers within the school. Six of these meetings were provided by external educational consultants who had also collaborated in the development of the PD programme. The remaining meetings were led by the school's internal project coach (see below). Various instructional formats were used, including interactive lectures and application of the strategies in practical exercises. Lesson Study (Murata, 2011) was also applied in adapted form: Teachers collectively prepared a mathematics lesson with specific attention for differentiation, one teacher taught the lesson and videotaped it, and the group evaluated the lesson afterwards.

The cycle of differentiation was the core of the PD programme. The materials for the PD programme were organised like a toolbox consisting of a Prezi presentation, practical application exercises, a Lesson Study guide, and literature about each step in the cycle of differentiation. To find a balance between specification of the programme and adaptivity to the needs of schools and teachers (Koellner & Jacobs, 2015), the educational consultants were asked to spend attention on each step of the cycle over the course of the year, but to select the most relevant topics based on the school's current needs. In the module about identification of educational needs, teachers learned how to use a broad range of formal and informal assessments to divide their students over flexible achievement groups and to determine the educational needs of their students both quantitatively and qualitatively. In the module about differentiated goals, teachers learned how to set differentiated goals for both high-achieving and low-achieving students. In the module about differentiated instruction, teachers learned how to provide a broad whole-class instruction, suitable for students of diverse achievement levels, as well as how to adapt subgroup instruction to the specific needs of low-achieving and high-achieving students. Central models were the levels of abstraction model and stages of learning mathematics model. Attention was also spent on prerequisite pedagogical content knowledge such as the order in which mathematical concepts are learned and knowledge about diverse solution procedures and typical mistakes. In the module about differentiated practice, teachers learned how to select the most important tasks for low-achieving students and how to apply compacting and enrichment for high-achieving students. In the module about evaluation, teachers learned how to use diverse formal and informal measures to assess students' progress towards the learning goals. They also learned how to reflect on the effectivity of the applied differentiation strategies to gain new information about students' educational needs.

Besides the team meetings for all teachers, the PD programme had two additional components: the instalment of internal project coaches and active involvement of the principal. At each school, at least two team members were trained to be a project coach, whose role was to function as a leader of change (Fullan, 2002). Project coaches led four of the team meetings – during which teaching teams discussed new schoolwide policies for differentiation and engaged in Lesson Study – and observed lessons of individual teachers to provide formative feedback about their application of differentiation. After the PD programme ended, project coaches were still available to coach and support their colleagues in further implementation of differentiation. Because of the importance of administrative support in the implementation of new educational practices (Klingner et al., 2003), principals were actively involved in an intake meeting and two intervision meetings to set goals, discuss progress and make plans to facilitate the implementation of differentiation.

3.2.4 Measures

To evaluate teachers' use of strategies for differentiation for students of diverse achievement levels in mathematics, the collected videos of mathematics lessons were scored with a standardised video observation instrument. Previously published observation instruments did not cover differentiation in sufficient depth for our purposes (e.g., the Mathematical Quality of Instruction instrument (MQI; Learning Mathematics for Teaching Project, 2011) or were not specific for mathematics and focused mainly on differentiation for advanced students (e.g., the Classroom Observation Scale – Revised (COS-R); VanTassel-Baska, Quek, & Feng, 2007) and the Differentiated Classroom Observation Scale (DCOS; Cassady et al., 2004). Therefore, we developed a new observation instrument called the Differentiation of Mathematical Instruction (DMI), inspired by the structure of the MQI (Learning Mathematics for Teaching Project, 2011).

The DMI consists of 16 items about differentiation for students of diverse achievement levels in mathematics. A list of items is provided in Table 3.3. The content of the items is based on expert consensus about best practice in differentiated mathematics education (see also Prast et al., 2015), supported by literature (e.g., Gersten et al., 2009; Rogers, 2007; Tomlinson et al., 2003; Van Groenestijn et al., 2011). Items 1 through 8 are evaluated

Table 3.3 DMI items

Item	Shortened item name
Whole-lesson items	
1. The teacher systematically works with three achievement groups (or more)	Achievement grouping
2. The teacher uses the levels of abstraction model	Levels of abstraction
3. The teacher uses the stages of learning mathematics model	Stages of learning
4. The teacher provides opportunities for student choice	Choice
5. The teacher and/or the students use multiple modalities in presenting or processing the lesson's content	Multiple modalities
6. Differentiation in practice tasks for low-achieving students	Differentiated practice LA
7. Differentiation in practice tasks for high-achieving students: compacting	Compacting HA
8. Differentiation in practice tasks for high-achieving students: enrichment	Enrichment HA
Fragment-specific items	
9. The teacher spends attention to low-achieving students	Attention LA
10. The teacher provides low-achieving students with the opportunity to work with manipulatives	Manipulatives LA
11. The teacher provides explicit instruction for low-achieving students	Explicit instruction LA
12. The teacher teaches at lower levels of abstraction for low-achieving students	Abstraction LA
13. The teacher spends attention to building understanding for low-achieving students	Building understanding LA
14. The teacher spends attention to high-achieving students	Attention HA
15. The teacher provides high-achieving students with challenging questions and tasks	Challenge HA
16. The teacher stimulates high-achieving students to reflect on the way of solving a problem	Reflection HA

LA = for / with low-achieving students, HA = for / with high-achieving students

for the lesson as a whole and concern general aspects of differentiation for students of diverse achievement levels with a focus on whether the strategies are used (rather than on the quality of differentiation). Items 9 through 16 consider adaptations to the specific needs of low-achieving and high-achieving students in a more detailed way. The lesson is split up in five-minute fragments and these items are scored separately for each fragment. All items are rated on a three-point scale. A score of 1 (low) indicates that the behaviour specified in the item is not or hardly present, a score of 3 (high) indicates that the behaviour is fully present for a substantial amount of time, and a score of 2 (mid) indicates that the behaviour is partially present but limited either in quality or in time. For each item, the instructional behaviour that needs to be observed to score low, mid or high is specified. Examples of this specification are provided in Appendix 3.1. The difference between item 8 and item 15 is that the whole-lesson item 8 simply evaluates whether the teacher uses enrichment tasks, regardless of the challenge level of these tasks (any task assigned specifically to high-achieving students is counted as an enrichment task) and regardless of the amount of time students spend working on these tasks. In contrast, item 15 evaluates whether high-achieving students are challenged in that specific fragment, either by challenging questions asked by the teacher or by enrichment tasks specifically developed for high-achieving students.

In addition to the items, the observer descriptively codes the instructional activities (e.g., whole-class instruction, subgroup instruction, etc.) in which students are engaged during at least half of the fragment (2.5 minutes). All fragment-specific items are scored for each fragment, regardless of the activity observed in that fragment. After coding, the scores on the fragment-specific items are averaged across the lesson. All items are averaged to compute a total score on the DMI (range 1.0 – 3.0).

To determine the interrater reliability of the DMI, a sample of 13 videos was coded independently by the first author and two trained research assistants holding a research master's degree in education. To assess interrater reliability, two-way mixed intraclass correlations (ICC's) with absolute agreement definition were calculated. Because each video in the whole dataset would be coded by one observer (contrary to this sample to assess the interrater reliability), the single-measures ICC was used. The ICC for the total score was .88, indicating high interrater reliability (Burdock, Fleiss, & Hardesty, 1963). Most single items also had a high interrater reliability (range .78 – .98). Two items had a somewhat lower ICC: use of the stages of learning mathematics model (item 3; ICC = .66) and providing opportunities for student choice (item 4; ICC = .62). These items were retained in the total score, but the item-level results for these two items should be interpreted with caution. Item 16 about stimulating reflection in high-achieving students had a low intraclass correlation (ICC = .27), probably because there was hardly any variance on this item[2]. Therefore, the item was removed from the total score, but it was retained in the item-level analyses since the low variance represents a finding in itself. After deletion of this item, Cronbach's alpha for the total score was .74 (based on the Year 1 pretest data).

3.2.5 Analyses

Item scores and total scores on the DMI were calculated per teacher per timepoint. When a teacher had been observed twice at one timepoint as planned, the scores of the two lessons were averaged. Otherwise, the scores of the one available observation were used. To explore which strategies were relatively (in)frequently used at baseline (research

[2] Stimulating reflection was almost never observed: in 97% of fragments all observers agreed on a score of 1, but in the remaining fragments observers disagreed on whether stimulating reflection was partly observable (score of 2) or not observable (score of 1).

question 1), the item-level means and medians were examined. To evaluate the effects of the PD programme on observed differentiation (research question 2), repeated-measures ANOVA's of the DMI total score with cohort as a between-groups factor were carried out. At all timepoints, the DMI total score was normally distributed and the variance was approximately equal across cohorts. Thus, the statistical assumptions for repeated-measures ANOVA were met. The data were analysed separately for each year because the sample of observed teachers partly differed between the years and because the number of experimental conditions varied between the years of the study: The Year 1 analyses compared the PD condition (Cohort 1) to the control condition (Cohort 2 and 3 together), whereas the Year 2 analyses distinguished between the PD (Cohort 2), follow-up (Cohort 1), and control (Cohort 3) condition. Finally, it was evaluated whether the cohorts differed significantly regarding the final implementation of the various strategies for differentiation (represented by single DMI items) at the posttest of Year 2, i.e., directly after the PD in Cohort 2 and one year after the ending of the PD in Cohort 1. Since some of the items had skewed or kurtose distributions (see Table 3.5), the nonparametric Kruskal-Wallis test followed up by Conover's pairwise comparisons (Conover, 1999) was used for these item-level analyses.

3.3 Results

3.3.1 Observed activities

A total of 406 lessons, with a mean duration of 50 minutes, was observed. Table 3.4 displays the activities in which students were engaged, split for low-achieving and high-achieving students[3]. For all students, the two most frequently observed activities were whole-class instruction and individual practice. The most important difference between low-achieving and high-achieving students was that low-achieving students frequently received additional subgroup or individual instruction (observed in 21% of fragments), whereas specific instruction for high-achieving students was very rare (2% of fragments). In 10% of fragments, some or all of the high-achieving students did not participate in the provided whole-class instruction and engaged in other activities – typically individual practice – instead. Accordingly, compared to average-achieving students, low-achieving students received more instruction and engaged less in individual practice whereas high-achieving students received less instruction and engaged more in individual practice.

[3] The activities of average-achieving students were not coded separately. Typically, average-achieving students mainly participated in whole-class instruction and independent practice activities, and occasionally joined a subgroup instruction or received individual instruction if necessary.

Table 3.4 Observed instructional activities of low-achieving (LA) and high-achieving (HA) students

	Percentage of fragments in which students were engaged in this activity[a]	
	LA	HA
Whole-class instruction	59.5%	53.4%
Subgroup instruction for LA / HA students	14.5%	1.4%
Individual instruction	6.5%	0.6%
Individual practice	54.6%	64.1%
Collaborative practice (heterogeneous groups)	5.1%	5.1%
Activities coded for HA students only		
Collaborative practice with other HA students	n/a	1.4%
Whole-class instruction is provided, but (some) HA students do not participate in it	n/a	10.0%

[a] During at least half of the fragment; percentages add up to more than 100%, because more than one activity could be coded during one fragment (subsequently or simultaneously).

3.3.2 Usage of differentiation strategies at baseline

At baseline, the mean total score on the DMI was 1.63 (*SD* = 0.24). Compared to the range of the instrument (from 1.00 to 3.00), this fell between low (score of 1 for an item in a fragment) and mid (score of 2). There was substantial variation between items, with mean scores ranging from 1.01 to 2.58. In Table 3.5, the items are rank-ordered based on their mean score[4]. Since the distribution of some items was skewed (see Table 3.5), the medians are also provided (ordering the items by their median would yield a similar rank order except that compacting would be ranked lower).

The most frequently observed item was achievement grouping. The majority of teachers (70.9%) obtained the maximum score of 3.0, indicating that they systematically worked with three achievement groups (regardless of the quality of instructional adaptations for these groups). Differentiation of the practice phase using enrichment tasks for high-achieving students and differentiated tasks for low-achieving students was also quite common with 40.0% and 30.9% of teachers obtaining the maximum score, respectively. However, compacting for high-achieving students was less frequently observed with only 5.5% of teachers obtaining the maximum score. Regarding general strategies for broadening instruction and practice to make it relevant for students of all achievement levels, the use of multiple modalities (*M* = 2.26) was more common than the use of the levels of abstraction model, the stages of learning mathematics model, and the provision

[4] Since Kruskall-Wallis tests did not reveal any significant differences between the cohorts at baseline, the item scores are displayed for all cohorts together.

of choice (range of means 1.46 – 1.71). The scores for the fragment-specific items about strategies to meet the specific needs of low-achieving and high-achieving students were generally quite low (range of means 1.01 – 1.43). This is partly a consequence of the structure of the instrument – e.g., teachers cannot score high on specific attention for low-achieving students in *all* fragments, since this would leave no time for attention for the other students in the class. However, there were still substantial differences between the various strategies. In line with the observed activities reported in section 3.1, specific attention for low-achieving students was much more common than specific attention for high-achieving students. Indeed, 72.7% of teachers scored 1.0 on attention for high-achieving students, indicating that they *never* spent specific instructional attention to these students for more than one minute at a stretch. Regarding qualitative adaptations for low-achieving students, explicit instruction, lower levels of abstraction, building understanding, and the use of manipulatives were all occasionally observed (range of means 1.29 – 1.43). For high-achieving students, only challenge (M = 1.32) was occasionally observed. In most cases, this challenge consisted of students working on pre-designed enrichment tasks (either provided in the regular curriculum or additional tasks designed specifically for high-achieving students) during part of the lesson, after finishing their regular curricular tasks. Stimulating reflection in high-achieving students was very seldomly observed, with only five teachers (9.1%) showing some limited application of this strategy.

Table 3.5 Ranking of DMI items from highest through lowest at pretest Year 1 (N = 55)

Item	Mean	*SD*	Median	Skewness[a]	Kurtosis[b]
1: Achievement grouping	2.58	0.73	3.00	-1.43*	0.45
8: Enrichment HA	2.27	0.72	2.50	-0.46	-1.10
5: Multiple modalities	2.26	0.59	2.00	-0.24	-0.73
6: Differentiated practice LA	2.02	0.76	2.00	0.13	-1.47
2: Levels of abstraction	1.71	0.71	1.50	0.53	-1.11
3: Stages of learning	1.56	0.71	1.50	0.71	-0.28
4: Choice	1.46	0.43	1.50	0.40	-1.04
7: Compacting HA	1.44	0.58	1.00	1.31*	1.00
11: Explicit instruction LA	1.43	0.29	1.38	0.48	-0.45
12: Abstraction LA	1.40	0.33	1.39	0.67	-0.03
9: Attention LA	1.36	0.29	1.40	0.63	0.01
13: Building understanding LA	1.36	0.31	1.30	1.00*	0.71
15: Challenge HA	1.32	0.28	1.30	0.60	-0.23
10: Manipulatives LA	1.29	0.37	1.13	1.24*	0.64
14: Attention HA	1.03	0.06	1.00	2.56*	6.56*
16: Reflection HA	1.01	0.03	1.00	3.24*	9.41*

LA = for / with low-achieving students, HA = for / with high-achieving students
[a] SE of skewness = 0.32 [b] SE of kurtosis = 0.63
* Value exceeds 3 times the standard error of the skewness (> 0.96) or kurtosis (> 1.89)

Taken together, the results indicate that most teachers already implemented some aspects of differentiation at baseline – especially achievement grouping combined with differentiated practice tasks as provided by the curriculum and additional instruction for low-achieving students – but that other, more qualitative and refined aspects of differentiation were used less frequently. Specific attention and qualitative adaptations seemed to be more commonly provided for low-achieving students than for high-achieving students.

3.3.3 Effects in Year 1

Table 3.6 reports the DMI total scores in Year 1, split by Cohort and timepoint. The repeated-measures ANOVA demonstrated no significant main effect of cohort (F (1, 53) = 0.20, p = .657, partial η^2 = .004) or timepoint (F (1, 53) = 0.91, p = .346, partial η^2 = .017). There was also no significant interaction between cohort and timepoint (F (1, 52) = 0.00, p = .999, partial η^2 = .000). In sum, experimental and control teachers showed similar levels of differentiation at both pretest and posttest.

Table 3.6 DMI total scores in Year 1

	Cohort 1 (PD) (*n* = 23)		Cohort 2 & 3 (control) (*n* = 32)	
	M	*SD*	*M*	*SD*
Pretest	1.62	0.26	1.64	0.22
Posttest	1.66	0.19	1.68	0.22

3.3.4 Effects in Year 2

DMI total scores for Year 2, split by Cohort and timepoint, are reported in Table 3.7. The repeated-measures ANOVA demonstrated a significant main effect of cohort (F (2, 56) = 4.81, p = .012, partial η^2 = .147). Following Cohen's (1988) guidelines (.01 ≈ small, .06 ≈ medium, .14 ≈ large), the magnitude of this effect can be classified as medium to large. Posthoc tests with Bonferroni correction showed that Cohort 1 scored significantly higher than both Cohort 2 (p = .040) and Cohort 3 (p =.035). Cohort 2 and 3 did not differ significantly from each other (p > .999). There was no main effect of timepoint (F (1, 56) = 0.86, p = .770, partial η^2 = .002). The interaction between cohort and timepoint did not reach significance (F (2, 56) = 1.43, p = .248, partial η^2 = .049). However, the differences between Cohort 1 versus Cohort 2 and 3 seemed to increase in the course of Year 2 because the scores of Cohort 1 showed an upward trend (d = +0.32) whereas the scores of Cohort 2 and 3 showed a downward trend (d = -0.21 for Cohort 2 and -0.25 for Cohort

Table 3.7 DMI total scores in Year 2

	Cohort 1 (post-PD) (*n* = 24)		Cohort 2 (PD) (*n* = 25)		Cohort 3 (control) (*n* = 10)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Pretest	1.55	0.16	1.49	0.21	1.46	0.15
Posttest	1.60	0.15	1.45	0.17	1.42	0.17

3). In sum, teachers of Cohort 1 (post-PD) scored significantly higher than teachers of Cohort 2 (PD) and Cohort 3 (control) throughout Year 2. Teachers of Cohort 2 and 3 did not differ significantly from each other.

3.3.5 Item-level differences at posttest Year 2

Table 3.8 displays the item-level scores at the posttest of Year 2 as well as the results of the Kruskal-Wallis test. The Kruskal-Wallis test revealed significant differences between the cohorts for item 5 (multiple modalities), 12 (abstraction LA), and 15 (challenge HA), and a marginally significant difference for item 4 (choice; $p = .050$). Conover's pairwise comparisons, which were used to follow up the Kruskal-Wallis test, indicated that teachers of Cohort 1 more often used multiple modalities (item 5), more often taught low-achieving students at a low level of abstraction (item 12), and more frequently provided choice (item 4) compared to teachers of Cohort 3. Teachers of Cohort 1 also provided more challenge to high-achieving students (item 15) compared to teachers of Cohort 2. The remaining pairwise comparisons were not significant ($p > .05$), indicating that Cohort 2 and 3 scored about equally on these items. For the remaining items, the Kruskal-Wallis test did not demonstrate significant differences, although Cohort 1 had the highest mean score for the majority of items.

Inspection of the mean item scores shows that there was still much variation between the different items at the posttest of Year 2. Across cohorts and similar to the baseline results, the highest scores were obtained for achievement grouping and differentiation of the practice tasks for high-achieving and low-achieving students, whereas the lowest scores were obtained for reflection and attention for high-achieving students. For reflection, there was no variance at all because all teachers scored 1.0. Regarding attention for high-achieving students it should be noted that, although the mean rank order did not significantly differ between the cohorts, the variance seems to have increased in Cohort 1 (Levene's test for equality of error variances: $F(2, 56) = 9.15$, $p < 0.001$). At the Year 2 posttest, eleven teachers of Cohort 1 (45.8%) spent at least some attention to high-achieving students with mean scores ranging from 1.04 to 1.39, compared to six teachers (24.0%) of Cohort 2 (range: 1.04 – 1.27) and three teachers of Cohort 3 (30.0%, range: 1.06 – 1.13).

Table 3.8 Item-level scores and differences between the cohorts at posttest Year 2

	Cohort 1 (n = 24)		Cohort 2 (n = 25)		Cohort 3 (n = 10)		Kruskal-Wallis	
Item	*M* (*SD*)	Median	*M* (*SD*)	Median	*M* (*SD*)	Median	*H* (*df* = 2)	*p*
1: Achievement grouping	2.60 (0.49)	3.00	2.44 (0.73)	3.00	2.50 (0.85)	3.00	0.42	.812
2: Levels of abstraction	1.31 (0.49)	1.00	1.24 (0.50)	1.00	1.00 (0.00)	1.00	4.20	.123
3: Stages of learning	1.19 (0.32)	1.00	1.28 (0.50)	1.00	1.15 (0.34)	1.00	0.65	.722
4: Choice	1.63 (0.50)	1.50	1.36 (0.42)	1.00	1.25 (0.35)	1.00	6.01	.050*[a]
5: Multiple modalities	1.73 (0.57)	1.50	1.54 (0.66)	1.50	1.20 (0.35)	1.00	6.96	.031*[a]
6: Differentiated practice LA	1.88 (0.73)	2.00	1.74 (0.54)	2.00	1.80 (0.89)	1.50	0.32	.851
7: Compacting HA	2.31 (0.79)	2.50	1.74 (0.81)	1.50	2.20 (0.89)	2.50	5.57	.062
8: Enrichment HA	2.40 (0.69)	3.00	2.40 (0.69)	2.50	2.35 (0.82)	2.75	1.99	.370
9: Attention LA	1.46 (0.33)	1.40	1.44 (0.32)	1.38	1.43 (0.30)	1.40	0.06	.968
10: Manipulatives LA	1.25 (0.37)	1.13	1.13 (0.23)	1.00	1.16 (0.43)	1.00	3.02	.221
11: Explicit instruction LA	1.54 (0.24)	1.54	1.56 (0.26)	1.53	1.45 (0.13)	1.41	2.06	.356
12: Abstraction LA	1.33 (0.28)	1.28	1.25 (0.23)	1.20	1.14 (0.24)	1.00	6.42	.040*[a]
13: Building understanding LA	1.35 (0.28)	1.27	1.32 (0.31)	1.25	1.27 (0.30)	1.23	0.71	.070
14: Attention HA	1.09 (0.13)	1.00	1.03 (0.07)	1.00	1.03 (0.05)	1.00	3.96	.138
15: Challenge HA	1.55 (0.38)	1.59	1.31 (0.27)	1.25	1.33 (0.31)	1.37	6.09	.048*[b]
16: Reflection HA	1.00 (0.00)	1.00	1.00 (0.00)	1.00	1.00 (0.00)	1.00	-[c]	-[c]

LA = for / with low-achieving students, HA = for / with high-achieving students
* $p \leq .05$ [a] Cohort 1 scores significantly higher than Cohort 3 ($p < .05$ on Conover's paired comparisons) [b] Cohort 1 scores significantly higher than Cohort 2
[c] No variation on this item

3.4 Discussion

This study about readiness-based differentiation in primary school mathematics used detailed video observations to answer the following two questions: (1) To what extent do teachers implement various strategies for differentiation at baseline, i.e., before specific PD? (2) What are the effects of a PD programme about differentiation in mathematics on teachers' implementation of differentiation?

3.4.1 Implementation of differentiation

At baseline, most teachers already implemented at least some aspects of differentiation, but some strategies were used more frequently than others. First, general structural aspects of differentiation seemed to be implemented more frequently than more refined, qualitative adaptations to the specific needs of low-achieving and high-achieving students. Many teachers used achievement grouping combined with differentiation of the practice tasks. These aspects of differentiation may be relatively easy to implement, especially if the mathematics curriculum provides ready-to-use tasks at three difficulty levels. Some general strategies to make whole-class instruction accessible for diverse students, including the use of multiple modalities, were also quite frequently observed. Such strategies may likewise be relatively easy to implement with little preparation time (Maeng & Bell, 2015). In contrast, more refined qualitative adaptations to the specific needs of low-achieving students (e.g., working with manipulatives) and high-achieving students (e.g., stimulating reflection) were observed less frequently. Working with such strategies may be more demanding because this requires advanced competencies including diagnostic competencies and knowledge of appropriate didactic strategies to meet the diagnosed educational needs.

Second, our data revealed differences in teachers' approach to differentiation for low-achieving versus high-achieving students. For low-achieving students, teachers focused on differentiation of instruction through additional subgroup or individual instruction. If the curriculum offered suggestions for differentiating the practice tasks for low-achieving students, teachers often followed these, but the changes compared to the regular practice tasks were often quite minor (e.g., a selection of the regular tasks). Conversely, for high-achieving students, the focus was on differentiation of the practice tasks rather than of the instruction. Many teachers used enrichment tasks, although the time working on these tasks was often limited since compacting was less frequently used (i.e., high-achieving students typically had to finish all regular tasks first). However, these enrichment tasks were seldomly discussed in subgroup or individual instruction for high-achieving students. This is problematic, because if high-achieving students are working on challenging enrichment tasks, they also need guidance and feedback. In fact, the challenge level of the tasks may be questioned if students never need feedback while working on them (VanTassel-Baska &

Stambaugh, 2005). The infrequent use of specific instructional attention for high-achieving students may also explain why teachers were almost never observed to stimulate reflection in high-achieving students. Teachers' tendency to allocate (additional) instructional time to low-achieving students rather than to high-achieving students may stem from a lack of awareness that high-achieving students also need guidance (Hertberg-Davis, 2009). An egalitarian culture which places more value on a sufficient achievement level of all students than on excellent achievement of some students may also contribute to this pattern. Indeed, international studies such as the Trends in International Mathematics and Science Study (TIMSS) have repeatedly shown that while almost all Dutch students reach at least a basic achievement level, relatively few Dutch students reach an excellent achievement level (i.e., high-achieving students in the Netherlands do not achieve as highly as their peers internationally; Meelissen & Punter, 2016).

3.4.2 Effects of the PD programme

The results regarding the effects of the PD programme are complex to interpret. In Year 1, scores were similar across timepoints and cohorts, indicating that the intervention had no observable short-term effects in Cohort 1. Similarly, there was no evidence for a positive short-term effect in Cohort 2 in Year 2. However, there was a main effect of cohort in Year 2, indicating that teachers from Cohort 1 applied more differentiation across the year. Notably, all cohorts seemed to show a drop in differentiation behaviour from the end of Year 1 to the beginning of Year 2. Although these scores cannot be directly compared because the sample of teachers (and their students) was partly different between years, we checked the results for teachers who were included in both years which yielded a similar pattern. This drop from the end of Year 1 to the beginning of Year 2 was present in all cohorts, although it seemed to be smaller in Cohort 1 than in Cohort 2 and 3. A potential explanation is that teachers usually get a new class of students at the beginning of the new schoolyear. Perhaps teachers need time to get to know their students before they can adapt instruction to their students' needs. The smaller drop in Cohort 1 might indicate that teachers who participated in the PD programme in Year 1, more swiftly started to diagnose and adapt to students' educational needs at the beginning of Year 2. The generally higher scores of Cohort 1 across Year 2 might be attributed to the intervention in Year 1. However, these findings should be interpreted cautiously. First, although the interaction between timepoint and cohort showed a trend in the expected direction, it was not significant. Second, as mentioned above, the sample was not identical across years. However, an additional analysis showed that teachers of Cohort 1 scored higher at the end of Year 2, even after controlling for differences at the beginning of the year (F (2,55) = 5.26, p = .008, partial η^2 = .16). This provides some support for the attribution of the effect to the intervention.

So, although our findings suggest that the PD programme may have had positive effects on the usage of differentiation strategies by teachers, they are not conclusive. The fact that there were no short-term effects in the year of intervention is in line with the study by VanTassel-Baska et al. (2008) in which effects started to emerge in year 2 and continued into year 3. However, the PD in that study continued for three years whereas the PD in the current study lasted only one year. If the main effect of Cohort 1 in Year 2 is interpreted as an effect of the intervention in Year 1, this would suggest that schools and teachers continued the process of implementing differentiation after the formal PD had ended. Indeed, some schools informed us that PD about differentiation was sustained in less formal ways, for example with continued Lesson Study cycles led by the project coach. Another important difference between the current study and the study by VanTassel-Baska et al. (2008) is that they used a standardised curriculum which provided teachers with much guidance regarding instructional adaptations. In contrast, teachers in our study were required to self-reliantly diagnose their students' needs and adapt instruction and practice using only general guidelines. A more standardised approach such as in the study by VanTassel-Baska et al. (2008) might not only make it easier to detect effects but also easier for teachers to implement differentiation.

At the posttest of Year 2, differences between the cohorts were most pronounced for teaching low-achieving students at a lower level of abstraction, challenging high-achieving students, using multiple modalities, and student choice. In all cases, teachers of Cohort 1 scored highest, although there was still room for improvement. Moreover, the differentiation strategies with the lowest mean scores at baseline – reflection and attention for high-achieving students – were still very infrequently used at the final posttest, also by teachers of Cohort 1. According to the trainers who provided the PD, the insight that high-achieving students also need specific guidance and feedback, especially when working on enrichment tasks, was an eye-opener for many participating teachers. Despite positive intentions to start providing such attention, teachers may have found it hard to reserve time for this during the mathematics lesson (Van de Weijer-Bergsma et al., 2016). Furthermore, previous research has indicated that insecurity about their own mathematical competence may make teachers hesitant to provide instruction about enrichment tasks to high-achieving students (Rubenstein, Gilson, Bruce-Davis, & Gubbins, 2015). Perhaps, developing the specific knowledge and skills to provide enrichment instruction to high-achieving students requires a more focused training than our rather broad PD programme. On a positive note, the observed increase in variance between teachers in Cohort 1 indicates that some teachers did start to spend more attention to high-achieving students.

3.4.3 Limitations, implications, and future research

The findings of the current study should be interpreted in light of the following limitations. First, the findings are based on a relatively small sample of teachers. Although this is common for video studies with intensive coding procedures, it may limit the generalisability of the findings. On the other hand, generalisability was positively influenced by the fact that data were collected at 31 schools spread across the country. A second limitation is that teachers with missing data at one timepoint per year were excluded from the study. Note, however, that missingness was mostly due to practical reasons (e.g., maternity leave, technical problems with the video) rather than due to teachers reclining further participation in the study, reducing the risk that these excluded teachers repesented a different category of teachers than those who were included. Third, a video observation is always a snapshot partly influenced by chance factors. To reduce this influence, we observed two lessons per teacher per timepoint, but when only one video was available, the teacher's score on one timepoint was based on only one lesson. This may have reduced the reliability and, accordingly, the power for detecting effects in the repeated-measures analyses. Thus, the results regarding the relative frequency of use of various differentiation strategies may be the most robust findings of this study, since these patterns were similar across timepoints and cohorts. A final limitation pertains to the observation instrument itself. While the development of a detailed instrument to measure differentiation in mathematics for students of all achievement levels is an important step forward, the downside of this innovative character is that the DMI has not been validated yet. Note, however, that interrater reliability was high. In future research, the DMI could be further developed and validated.

Despite these limitations, the present study adds to the literature by examining teachers' implementation of differentiation in primary school mathematics for students of all achievement levels using a detailed observation instrument. Our findings show that differentiation entails many different aspects, some of which are implemented more frequently than others. This may partly explain previous seemingly inconsistent findings regarding the implementation of differentiation and points to the need to specify explicitly which aspects of differentiation are studied in future research. Our current data demonstrate that, in line with previous self-report studies (Prast et al., 2015; Roy et al., 2013; Wan & Wan, 2013), teachers do implement many aspects of differentiation – especially the more structural ones including achievement grouping, differentiated tasks, and subgroup instruction for low-achieving students. At the same time, our data show that previously voiced concerns regarding teachers' implementation of differentiation (e.g., Hertberg-Davis, 2009; Inspectorate of Education, 2012; Schumm et al., 2000) may be warranted: The quality of differentiation, especially the use of more refined strategies to meet the educational needs of high-achieving and low-achieving students, could still be improved.

In light of the international trend towards inclusion of students with special educational needs, our findings imply that many teachers may not yet be fully prepared to provide these students with high-quality adaptations matched to their educational needs. Thus, this study not only supports the need for PD about achievement-based differentiation but also provides indications which aspects of differentiation may deserve most attention.

At the same time, this study shows that, in contrast to self-report studies which may find effects of brief PD (Edwards et al., 2006), even intensive and relatively long-term PD may yield only inconclusive evidence for observable effects on teachers' instructional behaviour. Previous observational studies have provided indications that PD programmes which are more focused and provide more guidance in the instructional materials may produce larger effects (Johnsen et al., 2002; VanTassel-Baska et al., 2008). This raises an interesting dilemma for future research: On the one hand, the nature of differentiation for students of diverse achievement levels seems to require a very broad set of knowledge and skills regarding both diagnosis of educational needs and instructional adaptations (Vogt & Rogalla, 2009). This requires teachers to see the bigger picture in which all components of the cycle of differentiation are interrelated (i.e., instructional adaptations should be based on diagnosis of educational needs and differentiated goals). This standpoint is in line with theories about adaptive teaching (Corno, 2008) which view the teacher as the expert in the learning process who should have the freedom to use the curriculum flexibly to adapt it to students' needs. Pre-differentiated curricula might hamper this flexibility. On the other hand, pre-differentiated curricula might make it easier and more feasible for teachers to implement differentiation since working with these might require less knowledge and preparation time. Previous research (Rubenstein et al., 2015) has shown that, with the help of carefully developed pre-differentiated curricula, many teachers were able to provide substantially differentiated lessons after only one day of PD. Our results also indicate that, if the curriculum provides ready-to-use tiered tasks, teachers tend to use these. We feel that pre-differentiated curricula are an interesting avenue for future research, if these curricula can in turn be used flexibly. For example, more suggestions for challenging questions and subgroup instruction for advanced students could be provided in the teaching manual. Furthermore, it is important that a good foundation of content knowledge and pedagogical content knowledge is laid in initial teacher education. This may apply especially to the domain of mathematics (Hill, Rowan, & Ball, 2005; Rubenstein et al., 2015). Given the major challenge of teaching students of diverse achievement levels, teachers deserve to be supported in multiple ways – including excellent intial teacher education, appropriate instructional materials, and opportunities for PD.

References

Brighton, C. M., Moon, T. R., & Huang, F. H. L. (2015). Advanced readers in reading first classrooms: Who was really "Left behind"? Considerations for the field of gifted education. *Journal for the Education of the Gifted*, *38*, 257-293. doi:10.1177/0162353215592501

Burdock, E. I., Fleiss, J. L., & Hardesty, A. S. (1963). A new view of inter-observer agreement. *Personnel Psychology*, *16*, 373-384. doi:10.1111/j.1744-6570.1963.tb01283.x

Cassady, J. C., Neumeister, K. L. S., Adams, C. M., Cross, T. L., Dixon, F. A., & Pierce, R. L. (2004). The differentiated classroom observation scale. *Roeper Review*, *26*, 139-146. doi:10.1080/02783190409554259

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York, NY: Wiley.

Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, *43*, 161-173. doi:10.1080/00461520802178466

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper-Perennial.

Deunk, M., Doolaard, S., Smale-Jacobse, A., & Bosker, R. J. (2015). *Differentiation within and across classrooms: A systematic review of studies into the cognitive effects of differentiation practices*. Groningen, The Netherlands: GION.

Dixon, F. A., Yssel, N., McConnell, J. M., & Hardin, T. (2014). Differentiated instruction, professional development, and teacher efficacy. *Journal for the Education of the Gifted*, *37*, 111-127. doi:10.1177/0162353214529042

Edwards, C. J., Carr, S., & Siegel, W. (2006). Influences of experiences and training on effective teaching practices to meet the needs of diverse learners in schools. *Education*, *126*, 580-592.

Fullan, M. (2002). The change leader. *Educational Leadership*, *59*(8), 16-21.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, *79*, 1202-1242. doi:10.3102/0034654309334431

Hertberg-Davis, H. (2009). Myth 7: Differentiation in the regular classroom is equivalent to gifted programs and is sufficient: Classroom teachers have the time, the skill, and the will to differentiate adequately. *Gifted Child Quarterly*, *53*, 251-253. doi:10.1177/0016986209346927

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*, 371-406. doi:10.3102/00028312042002371

Inspectorate of Education. (2012). *De staat van het onderwijs: Onderwijsverslag 2010-2011* [The state of education: Educational report 2010-2011]. Utrecht, The Netherlands: Inspectie van het Onderwijs.

Johnsen, S. K., Haensly, P. A., Ryser, G. A. R., & Ford, R. F. (2002). Changing general education classroom practices to adapt for gifted students. *Gifted Child Quarterly*, *46*, 45-63. doi:10.1177/001698620204600105

Klingner, J. K., Ahwee, S., Pilonieta, P., & Menendez, R. (2003). Barriers and facilitators in scaling up research-based practices. *Exceptional Children*, *69*, 411-429. doi:10.1177/001440290306900402

Koellner, K., & Jacobs, J. (2015). Distinguishing models of professional development: The case of an adaptive model's impact on teachers' knowledge, instruction, and student achievement. *Journal of Teacher Education*, *66*, 51-67. doi:10.1177/0022487114549599

Learning Mathematics for Teaching Project. (2011). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education*, *14*, 25-47. doi:10.1007/s10857-010-9140-1

Maeng, J. L., & Bell, R. L. (2015). Differentiating science instruction: Secondary science teachers' practices. *International Journal of Science Education*, *37*, 2065-2090. doi:10.1080/09500693.2015.1064553

McKenna, J. W., Shin, M., & Ciullo, S. (2015). Evaluating reading and mathematics instruction for students with learning disabilities: A synthesis of observation research. *Learning Disability Quarterly*, *38*, 195-207. doi:10.1177/0731948714564576

Meelissen, M. R. M., & Punter, R. A. (2016). *Twintig jaar TIMSS: Ontwikkelingen in leerlingprestaties in de exacte vakken in het basisonderwijs 1995 - 2015* [Twenty years of TIMSS: Developments in student achievement in mathematics and science in primary education 1995 - 2015]. Enschede, The Netherlands: Universiteit Twente.

Murata, A. (2011). Introduction: Conceptual overview of Lesson Study. In L. C. Hart, A. Alston, & A. Murata (Eds.), *Lesson Study research and practice in mathematics education* (pp. 1-12). London, UK: Springer.

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2015). Differentiation in primary school mathematics: Expert recommendations and teacher self-assessment. *Frontline Learning Research*, *3*, 90-116. doi:10.14786/flr.v3i2.163

Rogers, K. B. (2007). Lessons learned about educating the gifted and talented: A synthesis of the research on educational practice. *Gifted Child Quarterly*, *51*, 382-396. doi:10.1177/0016986207306324

Roy, A., Guay, F., & Valois, P. (2013). Teaching to address diverse learning needs: Development and validation of a differentiated instruction scale. *International Journal of Inclusive Education*, *17*, 1186-1204. doi:10.1080/13603116.2012.743604

Rubenstein, L. D., Gilson, C. M., Bruce-Davis, M. N., & Gubbins, E. J. (2015). Teachers' reactions to pre-differentiated and enriched mathematics curricula. *Journal for the Education of the Gifted*, *38*, 141-168. doi:10.1177/0162353215578280

Schumm, J. S., Moody, S. W., & Vaughn, S. (2000). Grouping for reading instruction: Does one size fit all? *Journal of Learning Disabilities*, *33*, 477-488. doi:10.1177/002221940003300508

Scott, B. J., Vitale, M. R., & Masten, W. G. (1998). Implementing instructional adaptations for students with disabilities in inclusive classrooms: A literature review. *Remedial and Special Education*, *19*, 106-119. doi:10.1177/074193259801900205

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*, 4-14. doi:10.3102/0013189X015002004

Smeets, E., Ledoux, G., Regtvoort, A., Felix, C., & Mol Lous, A. (2015). *Passende competenties voor passend onderwijs: Onderzoek naar competenties in het basisonderwijs* [Adaptive competencies for inclusive education: Research about competencies in primary school]. Nijmegen, The Netherlands: ITS.

Tieso, C. L. (2003). Ability grouping is not just tracking anymore. *Roeper Review*, *26*, 29-36. doi:10.1080/02783190309554236

Tieso, C. L. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted*, *29*, 60-89. doi:10.1177/016235320502900104

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., ... Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, *27*, 119-145. doi:10.1177/016235320302700203

Van de Weijer-Bergsma, E., Van Luit, J. E. H., Prast, E. J., Kroesbergen, E. H., Kaskens, J., Compagnie-Rietberg, C., … Logtenberg, H. (2016). *Differentiëren in het rekenonderwijs: Hoe doe je dat in de praktijk?* [Differentiation in mathematics education: How to do that in practice?]. Doetinchem, The Netherlands: Graviant scientific & educational books.

Van Groenestijn, M., Borghouts, C., & Janssen, C. (2011). *Protocol ernstige rekenwiskunde-problemen en dyscalculie* [Protocol severe mathematics difficulties and dyscalculia]. Assen, The Netherlands: Van Gorcum.

VanTassel-Baska, J., Feng, A. X., Brown, E., Bracken, B., Stambaugh, T., French, H., ... Bai, W. (2008). A study of differentiated instructional change over 3 years. *Gifted Child Quarterly*, *52*, 297-312. doi:10.1177/0016986208321809

VanTassel-Baska, J., Quek, C., & Feng, A. X. (2007). The development and use of a structured teacher observation scale to assess differentiated best practice. *Roeper Review*, *29*, 84-92. doi:10.1080/02783190709554391

VanTassel-Baska, J., & Stambaugh, T. (2005). Challenges and possibilities for serving gifted learners in the regular classroom. *Theory into Practice*, *44*, 211-217. doi:10.1207/s15430421tip4403_5

VanTassel-Baska, J., & Wood, S. (2010). The integrated curriculum model. *Learning and Individual Differences*, *20*, 345-357. doi:10.1016/j.lindif.2009.12.006

Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, *25*, 1051-1060. doi:10.1016/j.tate.2009.04.002

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wan, W. S., & Wan, W. E. (2013). Teachers' use of differentiation strategies in the Hong Kong classrooms. In E. H. F. Law, & C. Li (Eds.), *Curriculum innovations in changing societies: Chinese perspectives from Hong Kong, Taiwan and mainland China* (pp. 291-301). Rotterdam, The Netherlands: Sense Publishers.

Westberg, K. L., Archambault, F. X., Dobyns, S. M., & Salvin, T. J. (1993). The classroom practices observation study. *Journal for the Education of the Gifted*, *16*, 120-146. doi:10.1177/016235329301600204

Westberg, K. L., & Daoust, M. E. (2003). *The results of the replication of the classroom practices survey replication in two states*. Retrieved from http://www.gifted.uconn.edu/nrcgt/newsletter/fall03/fall03.pdf

Appendix 3.1
Sample specification of two DMI items

Sample specification of a whole-lesson item

1. The teacher systematically works with three achievement groups (or more)		
This item is about whether or not the teacher systematically and structurally works with achievement groups, rather than about the quality of differentiation for those groups.		
Low	Mid	High
No indications for working with achievement groups. Also score low if you only see some unplanned, ad hoc differentiation which is not combined with a relatively fixed grouping structure (e.g., individual instruction in response to the question of a student or spontaneous adaptation of practice tasks when some students finish early).	Some indications for working with achievement groups, but doubts whether a three-group structure is systematically used. E.g., only a subgroup instruction for low-achieving students (and no indications for use of a high-achieving subgroup).	Routine use of at least three achievement groups (low-achieving, average-achieving, and high-achieving; in multi-grade classes, if you see this in one grade you can assume that this also applies to the other grade within the class) which is indicated by for example: • relatively fixed groups (e.g., groups with names (e.g., the one-star group etc.), children know themselves to which group they belong, the seating arrangement corresponds with the groups) OR • use of a curricular method that prescribes differentiation in the practice tasks at three levels, and this differentiation is also applied (the teacher says which groups need to do which tasks or students know this themselves)

Sample specification of a fragment-specific item

9. The teacher spends attention to low-achieving students

This item is about attention of the teacher for low-achieving students in the form of subgroup instruction or individual instruction. Sometimes, it happens that the teacher starts a subgroup instruction, then tells the students to work on a practice task individually (still sitting at the instruction table, if applicable), leaves the students to work independently while he does something else, and resumes the subgroup instruction later. Therefore, it is the *interaction time* which counts: the time that the teacher actually interacts with the students (teacher talks or listens to students or actively watches how they are doing (for example by looking at their worksheets) and providing feedback if necessary). Examples include: instruction, guided practice, 'drill and practice' exercises carried out together (e.g., reciting multiplication tables together). To score individual instruction, the teacher should interact with the same student at least half of the fragment. So when the teacher provides 3 different students with a one-minute individual instruction, still score low (see 'high' for 1 exception).

Low	Mid	High
• There is no or hardly any (< 1 minute) subgroup instruction for low-achieving students AND • There is no individual instruction for low-achieving students or it lasts less than 2.5 minutes	• The teacher provides subgroup instruction to low-achieving students but the interaction time is short (1 – 2.5 minutes) *(Note: To distinguish specific attention for low-achieving students from ad hoc help in response to individual students' requests, short individual instruction does not count here)*	• The teacher provides subgroup instruction or individual instruction to low-achieving students and the interaction time is at least 2.5 minutes OR • The teacher provides two individual instructions to low-achieving students which each last approximately half of the fragment *(i.e., when both instructions are slightly shorter than 2.5 minutes, score high anyway)*


Hoeveel kost het?
blok 4, les 5
Hoeveel zit ertussen? Maak handig sprongen.

4

Differentiated instruction in primary mathematics: Effects of teacher professional development on student achievement

Emilie J. Prast
Eva Van de Weijer-Bergsma
Evelyn H. Kroesbergen
Johannes E. H. Van Luit

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2018). Differentiated instruction in primary mathematics: Effects of teacher professional development on student achievement. *Learning and Instruction*, *54*, 22-34. doi:10.1016/j.learninstruc.2018.01.009

Author contributions: All authors designed research; EvdW and EP collected the data under supervision of HvL and EK; EP analysed the data and wrote the paper; EvdW, EK and HvL provided feedback on the paper.

Abstract

This large-scale study examined the effects of a teacher professional development (PD) programme about differentiated instruction on students' mathematics achievement. Thirty primary schools (N = 5658 students of grade 1 – 6) divided over three cohorts participated: Cohort 1 received the PD programme in Year 1, Cohort 2 in Year 2, and Cohort 3 was control. During the PD, teachers learned how to adapt mathematics education to diverse educational needs using within-class ability groups. In Year 1, the PD had a significant small positive effect on student achievement growth. The effect size was similar for low-achieving, average-achieving and high-achieving students. In Year 2, no significant effects were demonstrated. In sum, teacher PD about differentiation has the potential to promote the achievement of all students. However, implementing differentiation is not straightforward and future research is necessary to unravel which factors make PD about differentiation succeed.

4.1 Introduction

Primary school classrooms are traditionally diverse in terms of the academic ability and achievement level of the students. With the current movement towards inclusion of children with special educational needs in general education classrooms, the range of ability and achievement levels is continuously increasing, as are the specific educational needs associated with these. Differentiation, i.e., the adaptation of instruction to students' different educational needs, is often promoted as a solution for responding to this diversity. In this study, we investigate whether teacher professional development (PD) about differentiation has a positive effect on student achievement in primary school mathematics.

4.1.1 Definitions: differentiation, ability grouping, and adaptive teaching competency

Roy, Guay, and Valois (2013, p.1187) define differentiated instruction as 'an approach by which teaching is varied and adapted to match students' abilities using systematic procedures for academic progress monitoring and data-based decision-making.' Thus, the focus is on differentiation based on students' current achievement level, also called cognitive or readiness-based differentiation. According to this definition, teachers should monitor students' academic progress to identify students' educational needs and then adapt instruction to these needs. The way in which progress is monitored and the nature of instructional adaptations can vary substantially, and various organisational formats can be used (e.g., individual or group-based; see Prast, Van de Weijer-Bergsma, Kroesbergen, & Van Luit (2015) for a discussion of this issue).

One frequently used way to organise differentiation is homogeneous within-class ability grouping (hereafter: ability grouping), in which students of similar academic ability or (current) achievement level are placed together in subgroups within the heterogeneous classroom (Tieso, 2003). Ability grouping is not synonymous to differentiation: it is an organisational format that can be used to implement differentiation, provided that instruction and practice are indeed adapted to the educational needs of the different ability groups.

A related term for adapting instruction to students' educational needs is adaptive teaching. A distinction is made between macro-adaptations (planned adaptations, e.g., pre-designed tasks at various levels of difficulty for low-achieving and high-achieving students) and micro-adaptations (spontaneous adaptations in direct response to students' needs; Corno, 2008). The term 'differentiation' seems to be more commonly used for macro-adaptations, whereas 'adaptive teaching' is more commonly used for micro-adaptations. However, the construct of 'adaptive teaching competency' (Vogt & Rogalla, 2009) does include both adaptive planning competency (teachers' capacity to plan adaptations beforehand; macro-adaptivity) and adaptive implementation competency (teachers' capacity for making adaptations on the spot; micro-adaptivity). In this article, we use 'differentiation' to refer to the process of monitoring progress and making instructional adaptations as defined by Roy et al. (2013). In line with Vogt & Rogalla (2009), we use 'adaptive teaching competency' to refer to teachers' capacities for making both planned and spontaneous adaptations to students' identified educational needs. We focus on planned adaptations based on students' current achievement level, but acknowledge that teachers should also be able to make adaptations on-the-fly in direct response to students' needs.

4.1.2 Achievement effects of ability grouping

Reviews about the effects of ability grouping have shown that positive effects can be obtained if instruction is tailored to the needs of the students in the subgroups and if the grouping arrangement is flexible (Kulik & Kulik, 1992; Lou et al., 1996; Slavin, 1987; Tieso, 2003). In contrast, slight negative effects of within-class ability grouping in primary school were found across three studies in which variations in instructional treatment were not explicitly described (Deunk, Doolaard, Smale-Jacobse, & Bosker, 2015).

An unresolved issue is the potential existence of differential effects depending upon achievement level. While Slavin (1987) reported a higher median effect size for low-achieving students than for average-achieving and high-achieving students, other reviews have found different patterns with smaller (Kulik & Kulik, 1992; Lou, Abrami, & Spence, 2000) or even negative effects (Deunk et al., 2015) for low-achieving students. Previously reported negative effects of ability grouping for low-achieving students have been ascribed to stigmatization and lower educational quality in low-ability groups (Gamoran, 1992). However, it has also been argued that these negative conditions can be prevented: negative stigma may be overcome by ensuring that the subgroups are within-class and flexible (Tieso, 2003) and by promoting a growth mindset rather than a fixed mindset of ability level (Dweck, 2000; i.e., participation in additional instruction should be communicated as an opportunity to learn, rather than as a sign of fixed low ability). Moreover, when ability grouping is used as a means to adapt education to the specific needs of the students in the groups, this may enhance (rather than reduce) educational quality for low-achieving students because the instruction can be better attuned to their needs (Gamoran, 1992). In an experimental study in which different types of ability grouping were compared and coupled with systematically prescribed instructional differentiation, Tieso (2005) found positive effects of flexible within-class grouping for all subgroups (low-achieving, average-achieving, and high-achieving).

4.1.3 Achievement effects of differentiation

A recent comprehensive literature review about the effects of differentiation on student achievement demonstrated that high-quality research about this topic is scarce (Deunk et al., 2015). For primary schools, only sixteen studies met the inclusion criteria, and most of these were still either too narrow (ability grouping only, without information about whether instructional adaptations were made; e.g., Leonard, 2001) or too broad (interventions in which differentiation was one of many components; e.g., Success for All; Borman et al., 2007) to specifically evaluate the effects of differentiation. However, promising findings were obtained with the five remaining studies, which demonstrated significant positive effects of two technological applications for differentiation. Individualizing Student Instruction (McDonald Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007; McDonald Connor et al., 2011a; McDonald Connor et al., 2011b) provides the teacher with recommendations about the amount and type of literacy instruction needed by individual students based on their scores on a computerised test. Accelerated Math (Ysseldyke et al., 2003; Ysseldyke & Bolt, 2007) continuously monitors students' progress and adapts practice tasks to students' individual skill level. While the review thus yielded evidence for the effectivity of technological applications for individual differentiation, studies in which (group-based) differentiation is mainly implemented by the teacher are scarce and often suffer from methodological limitations – most importantly small sample size and lack of a control group. Nevertheless, case studies of individual teachers and their classes (Brimijoin, 2002; Brown & Morris, 2005; Grimes & Stevens, 2009) do suggest that teachers may enhance the achievement of their students by implementing differentiation, although the generalisability of these findings may be limited due to the small sample size. In sum, there is some evidence to suggest that differentiation may promote student achievement in primary schools, especially when technological applications are used. However, there is still a need for large-scale studies in which differentiation is primarily in the hands of the teacher. While technological applications can be valuable for quantitative differentiation, teachers are still necessary for refined, qualitative diagnosis and adaptations.

4.1.4 Adaptive teaching competency

Teachers have an important role in enhancing student achievement: students of effective teachers achieve more (Nye, Konstantopoulos, & Hedges, 2004). According to the dynamic model of teacher effectiveness (Kyriakides, Creemers, & Antoniou, 2009), the most effective teachers distinguish themselves by the application of differentiation. Such teachers are skilled at adapting education to the needs of their students: they possess 'adaptive teaching competency' (Vogt & Rogalla, 2009). This requires extensive subject matter knowledge as well as advanced diagnostic, didactical, pedagogical, and classroom management skills (Smeets, Ledoux, Regtvoort, Felix, & Mol Lous, 2015; Vogt

& Rogalla, 2009). For teachers with less-developed knowledge and skills, implementing differentiation can be difficult. Many teachers feel that initial teacher education did not sufficiently prepare them for implementing differentiation (Inspectie van het Onderwijs, 2015). Therefore, a need for PD about differentiation has been identified (KNAW, 2009; Schram, Van der Meer, & Van Os, 2013).

4.1.5 Differentiation in mathematics using the cycle of differentiation

Against this background, project GROW (in Dutch, this is an acronym for differentiated mathematics education) was launched with the goal of developing and evaluating an effective PD programme for differentiation in primary school mathematics. We focused exclusively on mathematics, since domain-specific guidelines may provide teachers with more concrete advice for practical application than general guidelines. To ensure strong links between theory and practice, we collaborated intensively with a consortium of educational consultants and teacher trainers with expertise in mathematics. In the first stage of the project, we sought consensus among these experts about what teachers should do in daily practice to implement differentiation successfully. This resulted in the cycle of differentiation displayed in Figure 4.1 (see also Prast et al., 2015).

The cycle of differentiation starts with the identification of educational needs. First, the teacher should analyse the students' current skill level and divide the students over homogeneous achievement groups (typically low-achieving, average-achieving, and high-achieving). These achievement groups are used part of the time, besides whole-class instruction and individual practice and feedback, to cater specifically for the educational needs of the different subgroups. Students should be able to switch between groups based

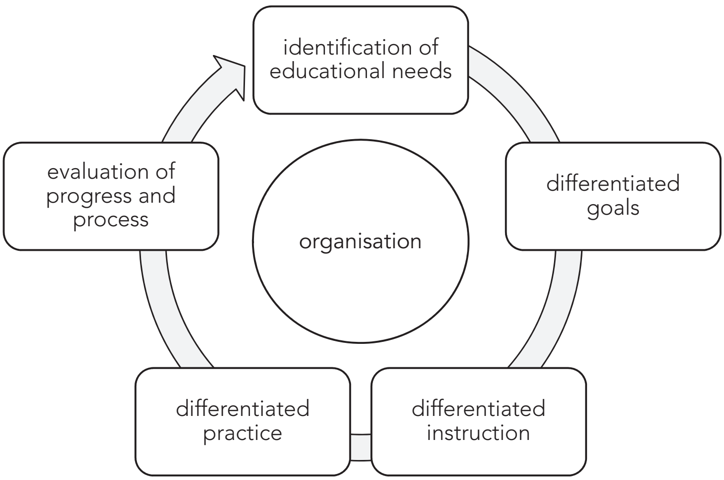


Figure 4.1 Cycle of differentiation (Prast et al., 2015; reprinted with permission).

on changes in their educational needs (Tieso, 2003). In addition to achievement tests, ongoing and refined diagnostic measures such as the analysis of daily work and diagnostic interviews should be used to signal changes in educational needs and to determine qualitative educational needs (i.e., why a student struggles with a particular type of sums and what the student needs to overcome this problem). In the second step, the teacher sets differentiated goals which should be challenging but realistic for the students in the different subgroups (Csikszentmihalyi, 1990). Third, the teacher differentiates instruction through broad whole-class instruction which engages students of diverse achievement levels, subgroup instruction tailored to the needs of that subgroup, and individual adaptations. One important way to differentiate instruction in mathematics is to use the stages from concrete to abstract mathematical reasoning (Gal'perin, 1969; Van Groenestijn, Borghouts, & Janssen, 2011). In subgroup instruction for low-achieving students, teachers need to spend more attention on concrete reasoning in order to build the understanding which underlies abstract reasoning. High-achieving students also need specific guidance and feedback, especially when they are working on appropriately challenging tasks (VanTassel-Baska & Stambaugh, 2005). Fourth, the practice tasks should be differentiated both quantitatively and qualitatively. For the low-achieving subgroup, the crucial tasks that are crucial for mastery of the goals for low-achieving students should be selected. For the high-achieving subgroup, the regular material should be compacted and enriched with challenging tasks which stimulate higher-level thinking (Rogers, 2007). Fifth, the teacher should evaluate whether the students have met the goals and whether the applied adaptations of instruction and practice had the desired effect using both formal (i.e., achievement tests) and informal measures (i.e., analysis of daily work). The evaluation phase informs the teacher about students' current achievement level and about instructional approaches that work for these students, completing the cycle and serving as new input for the identification of educational needs.

4.1.6 Research questions and hypotheses

The cycle of differentiation described above represents best practice as recommended by experts based on their experiential knowledge. However, as we have argued, quantitative empirical evidence proving that differentiation has positive effects on student achievement is scarce and there is still a need for large-scale studies in which differentiation is primarily in teachers' hands. In this article, we examine the effect of the PD programme developed for project GROW – in which teachers learn how to differentiate their mathematics lessons using the cycle of differentiation – on student achievement.

First, we investigate whether there is an overall effect of the PD programme on student achievement in the total sample. We expect a positive overall effect on achievement, because the PD programme should enable teachers to meet the educational needs of their students better.

Second, we examine whether the effects of the PD are similar or different for students of different achievement levels (differential effects). We hypothesise that the direction of effects is positive for students of all achievement levels, including low-achieving students. As we have argued, we expect that potential negative consequences of ability grouping for low-achieving students can be overcome by grouping students flexibly based on their current achievement level and by using this grouping structure to adapt education to the educational needs of the students (Gamoran, 1992; Slavin, 1987; Tieso, 2003). The PD programme should provide teachers with the knowledge and skills to make appropriate adaptations for students of diverse achievement levels, thereby using ability grouping as a means to differentiate instruction. Most previous reviews about within-class ability grouping have also yielded positive effect sizes for all achievement groups (Kulik & Kulik, 1992; Lou et al., 2000; Slavin, 1987), with exception of the review by Deunk et al. (2015) in which it was unclear whether and how the instruction was adapted to the needs of the students in the group. Besides the direction of the effects of the PD, we also explore whether the magnitude of effects differs between achievement groups (i.e., bigger or smaller effects for low-achieving or high-achieving students). Since previous reviews have been inconsistent about this (Deunk et al., 2015; Kulik & Kulik, 1992; Lou et al., 2000; Slavin, 1987), we do not formulate specific hypotheses regarding the relative magnitude of effects.

4.2 Method

4.2.1 Design

The design of the study is shown in Table 4.1. Participating schools were randomly assigned to one of three cohorts. In each cohort, data were collected across two schoolyears (i.e., all schools provided data on all measurement occasions), but the timing of the intervention differed between the cohorts: Cohort 1 participated in the PD programme in Year 1 and was a follow-up condition in Year 2, Cohort 2 was a control condition in Year 1 and participated in the PD programme in Year 2, and Cohort 3 served as a control condition in both years (but was offered to participate in the PD programme in the following schoolyear). Thus, we could examine the short-term effect of the intervention in two independent cohorts (Cohort 1 in Year 1 and Cohort 2 in Year 2) as well as the long-term effect (Cohort 1 in Year 2).

4.2.2 Participants

Schools were recruited with advertisements and flyers, with the proposed deal of free participation in the PD programme in combination with two years of data collection. Schools that were willing to participate could register themselves on a project website and we selected the first 32 schools that had registered. In the course of the project, two

Table 4.1 Research design

		Year 1 (2012–2013)			Year 2 (2013–2014)		
Cohort 1			PD programme			Follow-up	
Cohort 2			Control			PD programme	
Cohort 3			Control			Control	
Measurement occasions							
Mathematics test	T1		T2	T3		T4	T5
Nonverbal intelligence		a			b		
Visual-spatial working memory		a			b		
Verbal working memory			a		b		

Note. a = students in grade 1 – 6 in Year 1; b = students who enter grade 1 in Year 2.

of these schools dropped out. The first school (assigned to Cohort 1), dropped out after the first measurement occasion because it perceived the project as too intensive. The second school (assigned to Cohort 2), quit with the PD programme in the course of Year 2 after identifying other priorities for PD. Since the experimental condition of this school was neither purely control nor purely experimental, data collected at this school were disregarded. Thus, thirty schools spread across the Netherlands participated. These schools were diverse in terms of school size (M = 209 students per school, range 52 to 550) and mathematics curriculum used (five different curricula in different versions). Fifteen schools (50%) used single-grade classes. Nine schools (30%) used multi-grade classes (typically two adjacent grades within one classroom). Six schools (20%) used a combination of single-grade and multi-grade classes.

Data from all students in grade 1 through 6 were analysed (students who entered grade 1 in Year 2 only provided data in Year 2, students who left primary school in Year 2 only provided data in Year 1). The sample consisted of 196 classes in Year 1 and 186 classes in Year 2 (average class size: 24 students). In total, 5658 students (50.8% male) participated.

Table 4.2 provides descriptive information about the participating students and their teachers, split by year and cohort. In Year 1, student age differed significantly between the cohorts, $F(2, 4748) = 3.80$, $p = .023$, partial $\eta^2 = .002$. Pairwise comparisons indicated that students of Cohort 2 were significantly younger than students of Cohort 1 and 3 ($p < .05$ with Bonferroni correction). However, the effect size was very small and might be explained by students' grade level. That is, although grade levels were approximately equally represented in all cohorts (with 15.2 – 18.2% of students in each grade), Cohort 2 had relatively many students in grade 1 (18.2%) and relatively few students in grade 6 (15.2%) in Year 1. In Year 2, no age differences were found, $F(2, 4683) = 1.60$, $p = .202$, partial $\eta^2 = .001$. All subsequent analyses were controlled for grade level.

Table 4.2 Information about participants, split by Year and Cohort

	Cohort 1	Cohort 2	Cohort 3	Total
Students				
N Year 1	1514	1370	1867	4751
N Year 2	1494	1408	1790	4692
Age Year 1 (*M*, *SD*)	8.96 (1.82)	8.79 (1.82)	8.94 (1.86)	8.90 (1.84)
Age Year 2 (*M*, *SD*)	8.89 (1.80)	8.79 (1.83)	8.88 (1.83)	8.86 (1.82)
Gender Year 1 (% boys)	49.9%	50.3%	53.1%	51.3%
Gender Year 2 (% boys)	49.3%	49.6%	52.5%	50.6%
Teachers				
N Year 1	101	81	115	297
N Year 2	98	82	111	292
Years of experience Year 1 (*M*, *SD*)	16.54 (10.82)	13.17 (10.35)	14.80 (10.35)	15.11 (10.58)
Years of experience Year 2 (*M*, *SD*)	17.81 (10.91)	12.91 (9.49)	16.93 (10.82)	16.01 (10.75)
New at the school in Year 2 (*N*, %)	11 (11.2%)	13 (15.9%)	13 (11.7%)	37 (12.7%)

At the beginning of the study, teachers had an average of about fifteen years of teaching experience, with a broad range from zero to forty years. In Year 1, the mean number of years of experience of the teachers did not differ significantly across cohorts (F (2, 235) = 1.84, p = .160, partial η^2 = .016). In Year 2, teachers of Cohort 2 had significantly fewer years of experience than teachers of Cohort 1 and 3 (F (2, 242) = 4.73, p = .010, partial η^2 = .038; pairwise comparisons for Cohort 2 versus 1 and 3 were significant (p < .05 with Bonferroni correction)). The fact that this difference was only significant in Year 2 and not in Year 1 may be explained by the relatively large percentage of teachers who were new at the school in Year 2 in Cohort 2.

4.2.3 Measures

Mathematics achievement

Mathematics achievement was measured using the Cito Mathematics Tests (CMT; Janssen, Scheltens, & Kraemer, 2005a). These are national Dutch tests which are commonly administered at the middle and end of each schoolyear to monitor students' progress in mathematics throughout primary school. For each grade level, different versions with developmentally appropriate tasks for both the middle and end of the schoolyear have been developed (mid grade 1 through mid grade 6 – at the end of grade 6, a general end-of-primary-school test is nationally administered instead of the CMT). In all versions, five main domains are covered: (a) numbers and number relations, covering the structure of the number line and relations between numbers, (b) addition and subtraction, (c) multiplication and division, (d) complex math applications, often involving multiple mathematical manipulations, and (e) measuring (e.g., weight and length). From mid grade 2 to mid grade 6, the following domains are added successively: (f) estimation, (g) time, (h) money, (i) proportions, (j) fractions, and (k) percentages. The raw score on each grade-level test is converted into a mathematics competence score (for each raw score on each grade-level test, the CMT manual lists the corresponding competence score; thus, a competence score of 50 refers to the same competence level, regardless of which grade-level test was used). This competence score increases from grade 1 (minimum score: 0) through grade 6 (maximum score: 169) and can be used to assess growth in mathematics competence over time (Janssen, Scheltens, & Kraemer, 2005b). The reliability coefficients of the different versions range from .91 to .97 (Janssen, Verhelst, Engelen, & Scheltens, 2010). Based on a large sample which is representative for the Dutch population, norms are provided for each measurement occasion (Keuning et al., 2015). These include the mean competence score and its standard deviation for each grade level and timepoint (middle or end of the year).

Nonverbal intelligence

Since (nonverbal) intelligence has been shown to be an important predictor of mathematics achievement (Deary, Strand, Smith, & Fernandes, 2007; Geary, 2011), nonverbal intelligence was measured to be included in the model as a covariate. To this end, the Raven's Standard Progressive Matrices (SPM; Raven, Court, & Raven, 1996) was administered. Validity and reliability of the SPM as a measure of nonverbal, fluid intelligence are well-established (Schweizer, Goldhammer, Rauch, & Moosbrugger, 2007; Strauss, Sherman, & Spreen, 2006). Moreover, the Raven's SPM has demonstrated good internal consistency and predictive validity in the same sample as the current study (Van de Weijer-Bergsma, Kroesbergen, Jolani, & Van Luit, 2016).

The SPM consists of five series of 12 diagrams or designs with one part missing. Students should select the correct part which logically completes the designs. The difficulty level progressively increases over the test. A proportion correct score was calculated by dividing the total number of correct answers by the total number of items completed (students with missings on more than five items were treated as missing on the whole SPM). To control for the linear and quadratic effects of age, ageresidualised scores were created by regressing the proportion correct score on age and age-squared and saving the unstandardised residuals.

Working memory

Working memory – another important predictor of mathematics achievement (Friso-Van den Bos, Van der Ven, Kroesbergen, & Van Luit, 2013) – was also measured to be included as a covariate. Working memory was assessed with two online tasks suitable for self-reliant administration: the Lion game and the Monkey game. The Lion game is a visual–spatial complex span task (Van de Weijer-Bergsma, Kroesbergen, Prast, & Van Luit, 2015). Students are presented with a 4 × 4 matrix on the computer screen. In each trial, eight lions of different colours are consecutively presented at different locations in the matrix. Students have to remember the last location where a lion of a certain colour has appeared.

The Monkey game is a backward word span task (Van de Weijer-Bergsma et al., 2016). Students hear a number of spoken words, which they have to remember and recall backward by clicking on the words presented visually in a 3 × 3 matrix. For example, if students hear 'moon – fish – rose', they should click 'rose – fish – moon'. Both tasks consist of five levels in which working memory load is manipulated by increasing the number of lions or words (one through five) that students should remember. A mean proportion correct score indicating the proportion of lions or words recalled in the correct serial position was calculated and subsequently converted into an ageresidualised score.

Both tasks have demonstrated excellent internal consistency ($\alpha = .90$ for the Lion game and $\alpha = .87$ for the Monkey game) and have been shown to predict mathematics achievement ($\beta = .15$ for the Lion game and $\beta = .18$ for the Monkey game, $p < .001$) in the same sample as that of the current study (Van de Weijer-Bergsma et al., 2015; Van de Weijer-Bergsma et al., 2016). In addition, the Lion game has been shown to correlate ($r = .51$ - $.59$, $p < .001$) with the individually administered Automated Working Memory Assessment (Alloway, Gathercole, Kirkwood, & Elliott, 2008; Van de Weijer-Bergsma et al., 2015).

Evaluation questionnaire for teachers

At the end of the PD programme, teachers were asked to complete an evaluation questionnaire. In 15 items, teachers were asked to rate on a five-point Likert scale what they learned (based on the steps of the cycle of differentiation), whether they used what they learned in their daily mathematics teaching, and whether they perceived positive effects on their students' motivation and achievement. A sample item is: 'In the PD, I learned how to (better) diagnose my students' educational needs'. All items are provided in Table 4.3 (see section 4.3.1).

4.2.4 Procedure

Mathematics achievement was measured five times (see Table 4.1): at the middle and end of Year 1 and Year 2 and a baseline measurement at the end of the year before the study started (because the CMT is only administered at the middle and end of the schoolyear, it could not be administered at the beginning of Year 1). The CMT was administered by the classroom teacher. The SPM was group-administered in the classroom under supervision of a research assistant at the beginning of Year 1. A one-hour time limit was applied. The working memory tasks were administered online. Teachers were asked to make sure that their students completed the task self-reliantly within a specified time frame. The Lion game was administered at the beginning of Year 1. The Monkey game was still in development at that time so it was administered at the middle of Year 1. Students who entered grade 1 in Year 2 completed both working memory tasks and the SPM at the beginning of Year 2.

4.2.5 Professional development programme

Following the characteristics of effective teacher PD as summarised in a literature review by Borko, Jacobs, and Koellner (2010), the PD programme was designed to:

- connect to daily teaching practice and focus on students' learning
- include models of preferred instructional practice
- offer opportunities for active teacher learning
- stimulate collaboration and exchange between teachers
- offer multiple contexts, including classroom practice, for teacher learning
- be long-term, intensive and sustainable.

The PD programme consisted of three main components: PD for all teachers, an additional training for internal project coaches, and active involvement of the principal.

PD for all teachers

Ten three-hour team meetings spread across the schoolyear were provided for all teachers within the school. Six of these meetings were led by professional educational consultants who had collaborated in designing the PD programme as members of the consortium. The other four meetings were provided by the school's own project coaches (see below). During the team meetings, teachers learned about the cycle of differentiation and strategies for each step of the cycle. Attention was also spent on prerequisite knowledge, such as knowledge about the diverse solution procedures students use to solve particular types of problems and common mistakes. Various formats were used, including interactive lectures and application of the strategies in practical exercises. Lesson Study (Murata, 2011) was also applied in adapted form: Teachers collectively prepared a mathematics lesson with specific attention for differentiation, one teacher taught the lesson and videotaped it, and the group evaluated the lesson afterwards. Besides active participation in the team meetings, teachers were required to read selected literature and to apply certain strategies for differentiation in their mathematics lessons.

On the continuum from highly specified to highly adaptive approaches to PD (Koellner & Jacobs, 2015), we tried to find a balance between specification of the programme and adaptation to the needs and interests of specific schools and teachers. While the cycle of differentiation represented the common core of the PD programme, schools and teachers could also determine their own focus in consultation with the external educational consultant. To facilitate this adaptivity, the materials for the PD programme were organised like a toolbox consisting of a Prezi presentation, practical application exercises, and articles about the cycle of differentiation in general and practical strategies for each step. The educational consultants were asked to spend attention on each step of the cycle over the course of the year, but to select the most relevant exercises and literature based on the school's needs.

Project coaches

At each school, at least two team members were trained to be a project coach. The role of the project coach was to function as a change leader (Fullan, 2002) by coaching teachers in the process of implementing differentiated instruction. Project coaches were prepared for this role in five additional meetings which were organised regionally together with the project coaches of other participating schools. Meetings covered topics such as the analysis of the baseline situation and progress regarding differentiation within a school, the implementation of Lesson Study, and how to carry out classroom observations. Also, project coaches were required to read additional literature and write a paper about a self-selected aspect of differentiation relevant for their school. During the PD programme, project coaches gradually assumed more responsibility. Project coaches led four of the team meetings – during which teaching teams discussed new schoolwide policies for differentiation and engaged in Lesson Study – and observed lessons of individual teachers to provide formative feedback about their application of differentiation. After the PD programme ended, project coaches were still available to coach and support their colleagues in further implementation of differentiation. To enhance continued implementation, project coaches received a follow-up package which they could use for continued PD with the team and a train-the-trainer package to train new project coaches if necessary.

Involvement of the principal

Since administrative support is vital for successful implementation of new instructional practices (Klingner, Ahwee, Pilonieta, & Menendez, 2003), principals were actively involved. In an intake meeting, the educational consultant and the principal discussed the current situation in the school regarding differentiation and expectations about the PD programme. The roles and responsibilities of the principal, project coaches, and teachers were made explicit and attention was spent on how the principal could facilitate the PD programme. Principals were expected to be present at the team meetings. During the schoolyear, two

two-hour intervision meetings were organised for principals and project coaches to discuss progress, identify barriers to implementation, and make plans to facilitate implementation. Based on this, principals had to write a school-level plan for the continued implementation of differentiation in mathematics.

4.2.6 Analyses

The data were analysed with latent growth curve models using M*plus* version 7.31 (Muthén & Muthén, 1998–2012). First, the general effect of the intervention was evaluated. Subsequently, multiple-group models were used to evaluate whether the effect differed between achievement groups.

For the overall analysis, two models were estimated. Model 1 consisted of a latent growth curve model of mathematics achievement with control for covariates (see Figure 4.2). The analyses were carried out separately for each year of the study to enable separate evaluation of the effects in the intervention and post-intervention year[1]: The Year 1 model included T1, T2 and T3, with T2 specified as the intercept (since verbal working memory task – a predictor of the intercept – was administered at T2). The right-hand side of Figure 4.2 specifies the linear growth model for Year 1. The left-hand side of the figure lists the covariates, which were specified as predictors of the intercept and slope. The Year 2 model was analogous and included T3, T4 and T5 (T3 was used as the beginning point in this model because the CMT is not administered at the beginning of the schoolyear).

In Model 2, dummy variables representing the experimental conditions were added to the model to evaluate the effect of the PD programme. For the Year 1 analysis, the variable 'PD in Year 1' (coded as 1 for students in Cohort 1 and 0 for students in Cohort 2 and 3) was specified as an additional predictor of the intercept and slope to evaluate the short-term effect of the intervention on students in Cohort 1. For the Year 2 analysis, the variable 'PD in Year 2' (1 = Cohort 2, 0 = Cohort 1 and 3) was similarly added to evaluate the short-term effect of the intervention on students in Cohort 2. In addition, the variable 'PD in Year 1' was retained in the Year 2 analysis to evaluate the long-term effect of the PD on students in Cohort 1. In the interpretation of the results, we focus on the effect of the PD on the slope (rate of achievement growth). Effects on the intercept (level of achievement) are hard to interpret because the intercept is influenced by all timepoints in the model and, therefore, these analyses do not clarify whether any differences in level of achievement were already present at baseline or emerged over the course of the year as a result of the PD. Thus, the effect of the PD on the intercept was only included in the model to enable a purer evaluation of the effect of PD on the slope (controlling for any

[1] Rather than in a piecewise growth model for both years together, because the sample partly differed between years due to students entering grade 1 or leaving grade 6 and because students were not necessarily nested in the same classes in both years.

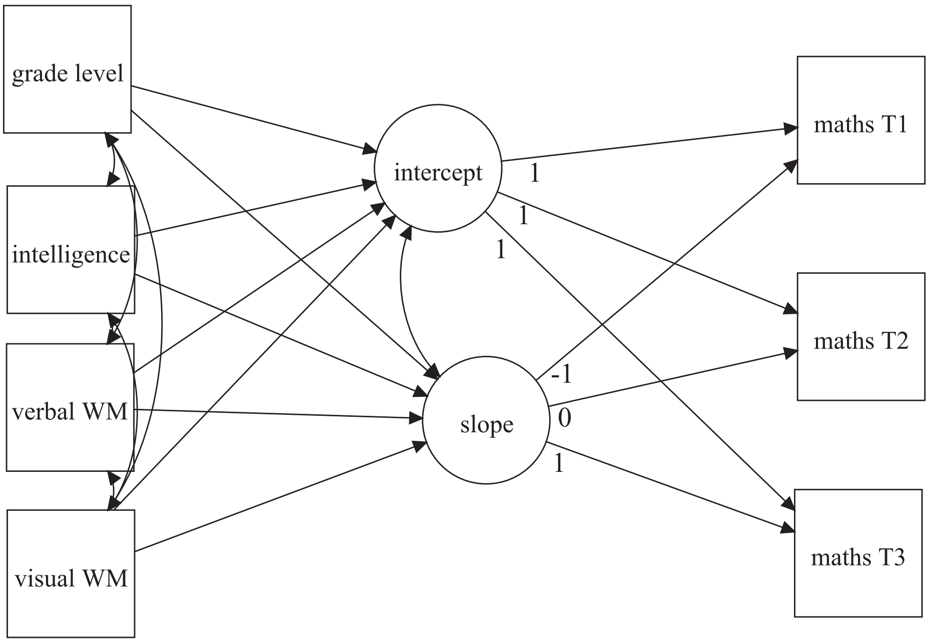


Figure 4.2 Model 1 for Year 1. WM = working memory.

differences between the cohorts in level of achievement) but the effect on the intercept itself was not interpreted. To test whether baseline mathematics achievement differed significantly between the cohorts, an additional ANOVA of the CMT scores at T1 with control for grade level was performed (see section 4.3.2).

Third, the full model (i.e., Model 2 from the overall analysis) was estimated as a multiple-group model for students of three achievement groups. Students were divided over three groups based on their CMT score at the first timepoint of the analysis (T1 for Year 1, T3 for Year 2). The multiple-group model was estimated for students of Grade 2 – 6 only, because students of Grade 1 had not yet entered primary school when this test was administered. Z-scores comparing students' competence score to the national norms (*M* and *SD* on each grade-level test) were computed. To create three approximately equally sized groups, students with z-scores below -0.5 were assigned to the low-achieving group, z-scores above 0.5 to the high-achieving group, and z-scores between -0.5 and +0.5 to the average-achieving group. Wald tests were used to evaluate whether the parameters estimating the effect of the PD on the intercept and slope were significantly different between achievement groups, which would be an indication of differential effects.

In all analyses, the 'type=complex' option in M*plus* was used to control for the nesting of students within classes. This method ensures that standard errors are corrected for the clustered data structure without building a full multilevel model (McNeish, Silverman, & Stapleton, 2017). In our case, multilevel modelling was complicated since grade level was neither purely an individual-level variable nor purely a class-level variable due to the existence of multigrade classes. Single-level analysis methods with cluster-robust standard errors (such as type=complex) are an appropriate and computationally less demanding alternative for multilevel modelling (McNeish et al., 2017). Model fit was evaluated using the chi-square statistic, the comparative fit index (*CFI*), the Tucker-Lewis Index (*TLI*), the root mean squared error of approximation (*RMSEA*), and the standardised root mean square residual (*SRMR*). Due to the large sample size, the chi-square statistic was expected to be significant. The models were judged to have a good fit if they had values above .95 for the *CFI* and *TLI* and values below .06 and .08 for the *RMSEA* and *SRMR*, respectively (Hu & Bentler, 1999).

4.3 Results

4.3.1 Teacher participation in and evaluation of the PD

In Year 1, 81 teachers of Cohort 1 (81.0%) obtained their certificate for participation, indicating presence at least eight out of ten team meetings. In Year 2, 72 teachers of Cohort 2 (90%) obtained their certificate. Although reasons for absence were not always known to us, teachers who missed many team meetings often had reasons such as having left or entered the school in the course of the year, long-term illness, maternity leave, or a part-time job (i.e., teachers were asked to attend the team meetings that were planned on days they did not work, but this was not always possible).

The teacher evaluation questionnaire about the PD programme was completed by 76 teachers of Cohort 1 at the end of Year 1 and 73 teachers of Cohort 2 at the end of Year 2. As can be seen in Table 4.3, teachers were moderately positive about what they learned in the PD, with scores above the midpoint of the scale for all questions. Teachers indicated that they had learned about all steps in the cycle of differentiation. Moreover, the majority of teachers (76.3% and 76.7% of teachers who completed the questionnaire in Cohort 1 and 2, respectively) mostly or fully agreed that they actually used what they had learned in the PD for the preparation and teaching of their mathematics lessons. Teachers also perceived positive effects of implementing (more) differentiation on students' motivation and achievement.

Table 4.3 Evaluation of the PD by participating teachers

	Cohort 1 (Year 1)		Cohort 2 (Year 2)	
	M	*SD*	*M*	*SD*
In the PD, I extended my knowledge about mathematics education in general (e.g., didactics)	3.69	0.78	3.56	0.97
In the PD, I learned how to (better) …				
… Diagnose my students' educational needs	3.66	0.76	3.47	0.86
… Set differentiated goals	3.78	0.72	3.34	0.98
… Broaden whole-class instruction	3.35	0.89	3.17	1.00
… Adapt instruction for low-achieving students	3.17	0.97	3.50	1.05
… Adapt practice for low-achieving students	3.15	0.82	3.27	0.96
… Adapt instruction for high-achieving students	3.56	0.80	3.36	0.99
… Adapt practice for high-achieving students	3.58	0.87	3.23	0.99
… Evaluate whether my chosen way of teaching was effective for my students	3.15	0.86	3.09	1.00
… Organise differentiation in practice (e.g., working with subgroups)	3.33	0.99	3.29	1.13
… Apply (more) differentiation in my mathematics lessons	3.53	0.76	3.36	1.05
I can use what I learned in the PD for the preparation and teaching of my mathematics lessons	3.84	0.84	4.00	0.79
I actually use what I learned in the PD for the preparation and teaching of my mathematics lessons	3.90	0.78	3.86	0.87
Applying (more) differentiation in my mathematics lessons has a positive effect on my students' motivation	3.87	0.73	3.70	0.83
Applying (more) differentiation in my mathematics lessons has a positive effect on my students' achievement	3.47	0.78	3.57	0.69

Note. 1 = fully disagree, 5 = fully agree.

4.3.2 Descriptive statistics and missing data

Descriptive statistics of students' scores on the mathematics tests, the nonverbal intelligence test, and the two working memory measures are displayed in Table 4.4. An ANOVA comparing the raw competence scores of the cohorts on the mathematics test at T1 showed a significant but very small effect of cohort (F (2, 3511) = 3.15, p = .043, partial η^2 = .002; pairwise comparisons not significant). However, after controlling for grade level, these differences disappeared (F (2, 3510) = 1.80, p = .165, partial η^2 = .002; pairwise comparisons not significant). Thus, students of Cohort 1 (estimated mean[2] = 77.88, *SE* = 0.41), Cohort 2 (estimated mean = 77.41, *SE* = 0.43), and Cohort 3 (estimated mean = 76.84, *SE* = 0.37) had comparable baseline scores.

[2] After controlling for grade level; raw means and standard deviations are reported in Table 4.4.

Table 4.4 Descriptive statistics

	Cohort 1	Cohort 2	Cohort 3	Total			
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*Min.*	*Max.*	*n* (%)
Maths T1	76.78 (24.80)	76.18 (24.46)	78.59 (24.96)	77.33 (24.78)	7.00	143.00	3514 (73.96)[b]
Maths T2	77.34 (30.87)	74.87 (29.77)	76.90 (29.90)	76.45 (30.18)	0.00	154.00	4448 (93.62)[b]
Maths T3	77.95 (25.33)	74.39 (25.99)	74.24 (25.50)	75.48 (25.65)	0.00	149.00	3523 (74.06)[b,c]
Maths T4	75.13 (28.92)	75.20 (29.11)	75.91 (28.70)	75.44 (28.89)	0.00	154.00	4048 (86.27)[c]
Maths T5	78.89 (28.92)	77.43 (24.72)	77.95 (24.67)	78.06 (24.44)	0.00	164.00	3358 (71.57)[c]
Nonverbal intelligence[a]	-0.78 (7.47)	0.21 (7.76)	0.36 (7.75)	-0.04 (7.68)	-30.36	23.16	4998 (88.33)[d]
Verbal WM[a]	0.00 (0.14)	0.01 (0.14)	-0.01 (0.15)	0.00 (0.14)	-0.54	0.44	4618 (81.62)[d]
Visual-spatial WM[a]	0.01 (0.15)	0.00 (0.17)	-0.01 (0.17)	0.00 (0.16)	-0.72	0.41	4763 (84.18)[d]

Note. WM = working memory
[a] ageresidualised score [b] percentage of students in Year 1 (N = 4751) [c] percentage of students in Year 2 (N = 4692) [d] percentage of total number of students (N = 5658)

Grade level was uniformly distributed with approximately 17% of students in each grade level. The other variables approximated the normal distribution, but some skewness and kurtosis was present. Therefore, the Maximum Likelihood Robust estimator, which is robust to deviations from normality, was used in all subsequent analyses.

The percentage of available data – and, conversely, the percentage of missing data – is provided in the last column of Table 4.4. Most of the missing data on the mathematics test are missing by design because the CMT is neither administered before the start of grade 1 nor at the end of grade 6. Remaining causes for missingness are absence on the day of testing and – in case of the working memory tasks – technical problems with the games and lack of systematic administration by some teachers. M*plus* can handle missing data well by making flexible use of all relevant available information for each parameter. To enable the inclusion of cases with missing values on one or more covariates (which are, by default, completely removed from the analysis in M*plus*), we specified the variances of the covariates as parameters to be estimated in all models.

4.3.3 Overall analysis Year 1

Model 1 had a good fit: *RMSEA* = .024, *CFI* = 0.999, *TLI* = 0.998, *SRMR* = .012. As expected, the chi-square test was significant: χ^2 (5) = 18.13, *p* = .003. The growth model explained over 95% of the variance in the observed variables. Model results are displayed in Table 4.5. Regarding the prediction of the latent variables, all covariates had a significant positive effect on the intercept and this effect was largest for grade level. Only grade level had a significant effect on the slope, and this effect was negative (i.e., students in lower grade levels acquired new knowledge and skills at a faster pace). Taken together, grade level, nonverbal intelligence, visual-spatial working memory and verbal working memory explained 88% of the variance of the intercept and 16% of the variance of the slope of mathematics achievement.

Model 2, in which the effect of the intervention was added, had a good fit: χ^2 (10) = 19.89, *p* = .030, *RMSEA* = .014, *CFI* = 1.000, *TLI* = 0.999, *SRMR* = .017. PD in Year 1 had a significant but small positive effect on the slope: β = 0.15, *p* = .007. Thus, students in Cohort 1 gained about 2.5 points *more* on the CMT in the course of Year 1 than students in the other cohorts (average growth is 14.4 points). Adding the effect of the intervention to the model explained an additional 2% of the slope variance. In sum, in line with our hypothesis, the intervention had a positive short-term effect on the slope of mathematics achievement in Cohort 1.

4.3.4 Multiple-group model Year 1

The multiple-group model, in which the full model was estimated separately for three achievement groups, initially yielded two negatively estimated residual variances (for mathematics T1 and T3) in the average-achieving group. This problem was solved by

Table 4.5 Overall model Year 1 ($N = 4751$)

	Model 1			Model 2		
Parameter	Estimate	*SE*	*p*	Estimate	*SE*	*p*
Predictors of the intercept[a]						
Grade level	0.87	0.01	< .001	0.87	0.01	< .001
Nonverbal intelligence	0.20	0.01	< .001	0.20	0.01	< .001
Verbal WM	0.09	0.01	< .001	0.09	0.01	< .001
Visual-spatial WM	0.07	0.01	< .001	0.07	0.01	< .001
PD in Year 1	n/a			0.04	0.01	< .001
Predictors of the slope[a]						
Grade level	-0.40	0.08	< .001	-0.40	0.08	< .001
Nonverbal intelligence	-0.03	0.03	.338	-0.02	0.03	.600
Verbal WM	0.07	0.04	.069	0.07	0.04	.082
Visual-spatial WM	0.01	0.03	.719	0.00	0.03	.992
PD in Year 1	n/a			0.15	0.05	.007
Correlations[a]						
Intercept with slope	0.00	0.05	.937	-0.02	0.05	.744
Nonverbal intelligence with grade level	0.07	0.02	.004	0.07	0.02	.004
Verbal WM with grade level	0.08	0.03	.001	0.08	0.03	.001
Verbal WM with nonverbal intelligence	0.40	0.02	< .001	0.40	0.02	< .001
Visual-spatial WM with grade level	0.08	0.02	.001	0.08	0.02	.001
Visual-spatial WM with nonverbal intelligence	0.37	0.02	< .001	0.37	0.02	< .001
Visual-spatial WM with verbal WM	0.35	0.02	< .001	0.35	0.02	< .001
Intercepts[b]						
Intercept	76.54	0.33	< .001	75.78	0.38	< .001
Slope	7.53	0.18	< .001	7.21	0.22	< .001
Residual variances[b]						
Maths T1	28.34	5.22	< .001	28.50	5.15	< .001
Maths T2	42.82	4.66	< .001	41.75	2.61	< .001
Maths T3	28.82	4.66	< .001	29.32	4.41	< .001
Intercept	104.35	3.49	< .001	103.15	3.37	< .001
Slope	8.50	2.23	< .001	8.20	2.16	< .001
Explained variances						
Maths T1	0.97	0.01	< .001	0.97	0.01	< .001
Maths T2	0.95	0.00	< .001	0.95	0.00	< .001
Maths T3	0.97	0.01	< .001	0.97	0.01	< .001
Intercept	0.88	0.01	< .001	0.88	0.01	< .001
Slope	0.16	0.06	.015	0.18	0.06	.006

Note. WM = working memory. For parsimony, the means (all close to 0 due to centering) and variances of the covariates are omitted from the table.
[a] standardised [b] unstandardised

fixing the residual variance of mathematics T1 to 0 in this group. This solution was deemed acceptable, since the model generally explained a very large proportion of the variance in the observed mathematics scores (leaving little residual variance) and since the variance was likely to be smaller within the groups because they were created based on mathematics achievement at T1. After fixing this residual variance to 0, the model had a good fit: χ^2 (31) = 102.89, *p* < .001, *RMSEA* = .044, *CFI* = 0.997, *TLI* = 0.995, *SRMR* = .035. As can be seen in Table 4.6, the results were largely similar to the overall model, although the effects of the covariates and their correlations differed somewhat. In addition to the previously found effects, nonverbal intelligence had a significant positive effect on the slope within all achievement groups and verbal working memory had a significant positive effect on the slope within the average-achieving and high-achieving group.

PD in Year 1 had a significant positive effect on the slope of mathematics achievement for average-achieving students (β = 0.10, *p* = .040) and high-achieving students (β = 0.12, *p* = .036). For low-achieving students, the effect was similar in size but did not reach significance (β = 0.12, *p* = 0.051). However, Wald tests demonstrated that the effect of PD on achievement growth was not significantly different between achievement groups (low-achieving vs. average-achieving students: *W* = 0.19, *p* = .667; low-achieving vs. high-achieving: *W* = 0.01, *p* = .913; average-achieving vs. high-achieving: *W* = 0.08, *p* = 0.776). Thus, the fact that the effect did not reach significance in the low-achieving group probably does not reflect a different effect size but may be a consequence of the slightly smaller sample size of the low-achieving subsample. Therefore we conclude that, in Year 1, the intervention had a positive effect on mathematics achievement growth for all achievement groups, in line with our hypothesis. Since these effects were similar across achievement groups, we found no evidence for differential effects.

4.3.5 Overall analysis Year 2

Model 1 of the Year 2 analysis had a good fit: χ^2 (5) = 37.29, *p* < .001, *RMSEA* = .037, *CFI* = 0.999, *TLI* = 0.996, *SRMR* = .015. As can be seen in Table 4.7, the results of Model 1 in Year 2 resembled the results of Model 1 in Year 1. The effects of the covariates on the intercept and slope were similar and, taken together, explained 88% of the intercept variance and 22% of the slope variance. The fit of Model 2 was good as well: χ^2 (15) = 23.66, *p* = .071, *RMSEA* = .011, *CFI* = 1.000, *TLI* = 1.000, *SRMR* = .017. However, adding the effect of the intervention did not explain additional variance. In contrast to the Year 1 findings, participation in the PD programme in Year 2 did not have a significant short-term effect on the slope (β = 0.03, *p* = .640). Regarding the long-term effect of the intervention, PD in Year 1 had no significant effect on the slope of students in Cohort 1 in Year 2 (β = -0.06, *p* = .665). In sum, in contrast to our hypothesis, neither short-term nor long-term effects of the intervention on mathematics achievement growth could be demonstrated in Year 2.

Table 4.6 Multiple-group model Year 1

	Low-achieving *n* = 989			Average-achieving *n* = 1300			High-achieving *n* = 1225		
Parameter	Estimate	*SE*	*p*	Estimate	*SE*	*p*	Estimate	*SE*	*p*
Predictors of the intercept[a]									
Grade level	0.92	0.01	< .001	0.96	0.00	< .001	0.93	0.01	< .001
Intelligence	0.09	0.01	< .001	0.05	0.01	< .001	0.11	0.01	< .001
Verbal WM	0.06	0.02	< .001	0.03	0.01	< .001	0.06	0.01	< .001
Visual-spatial WM	0.04	0.02	.010	0.02	0.01	.083	0.03	0.01	.021
PD in Year 1	0.04	0.01	.002	0.02	0.01	.052	0.05	0.01	< .001
Predictors of the slope[a]									
Grade level	-0.34	0.07	< .001	-0.29	0.05	< .001	-0.24	0.07	< .001
Intelligence	0.15	0.05	.003	0.13	0.04	.001	0.12	0.05	.007
Verbal WM	0.07	0.04	.099	0.11	0.04	.001	0.14	0.05	.003
Visual-spatial WM	0.08	0.05	.161	0.06	0.04	.084	-0.06	0.05	.197
PD in Year 1	0.12	0.06	.051	0.10	0.05	.040	0.12	0.06	.036
Correlations[a]									
Intercept with slope	0.45	0.09	< .001	0.62	0.03	< .001	0.15	0.08	.048
Intelligence with grade level	0.08	0.04	.052	0.06	0.04	.075	0.03	0.04	.379
Verbal WM with grade level	0.07	0.04	.083	0.07	0.04	.063	0.03	0.04	.524
Verbal WM with intelligence	0.26	0.03	< .001	0.27	0.03	< .001	0.33	0.03	< .001
Visual-spatial WM with grade level	0.16	0.04	< .001	0.03	0.03	.063	0.01	0.05	.923
Visual-spatial WM with intelligence	0.38	0.04	< .001	0.23	0.03	< .001	0.19	0.03	< .001
Visual-spatial WM with verbal WM	0.31	0.04	< .001	0.23	0.03	< .001	0.24	0.04	< .001

Table 4.6 continues on next page

Table 4.6 *Continued*

Parameter	Low-achieving *n* = 989			Average-achieving *n* = 1300			High-achieving *n* = 1225		
	Estimate	*SE*	*p*	Estimate	*SE*	*p*	Estimate	*SE*	*p*
Intercepts[b]									
Intercept	69.63	0.44	< .001	84.31	0.27	< .001	95.12	0.35	< .001
Slope	8.17	0.29	< .001	7.24	0.23	< .001	5.36	0.29	< .001
Residual variances[b]									
Maths T1	17.00	6.31	.007	0.00[c]	n/a	n/a	17.00	6.31	.007
Maths T2	55.21	4.27	< .001	40.49	3.16	< .001	55.21	4.27	< .001
Maths T3	23.13	6.72	< .001	0.67	4.15	.872	23.13	6.72	.001
Intercept	41.14	3.05	< .001	26.16	1.56	< .001	41.14	3.05	< .001
Slope	13.97	3.16	< .001	15.40	1.43	< .001	13.97	3.16	< .001
Explained variance									
Maths T1	0.98	0.01	< .001	1.00	n/a	n/a	0.96	0.01	< .001
Maths T2	0.92	0.01	< .001	0.91	0.01	< .001	0.88	0.01	< .001
Maths T3	0.99	0.02	< .001	0.99	0.01	< .001	0.95	0.02	< .001
Intercept	0.90	0.01	< .001	0.94	0.01	< .001	0.90	0.01	< .001
Slope	0.16	0.04	< .001	0.13	0.03	< .001	0.11	0.04	.004

Note. WM = working memory, intelligence = nonverbal intelligence. For parsimony, the means (all close to 0 due to centering) and variances of the covariates are omitted from the table.
[a] standardised [b] unstandardised [c] fixed to 0

Table 4.7 Overall model Year 2 (*N* = 4692)

	Model 1			Model 2		
Parameter	Estimate	*SE*	*p*	Estimate	*SE*	*p*
Predictors of the intercept[a]						
Grade level	0.87	0.01	< .001	0.87	0.01	< .001
Intelligence	0.19	0.01	< .001	0.20	0.01	< .001
Verbal WM	0.10	0.01	< .001	0.10	0.01	< .001
Visual-spatial WM	0.08	0.01	< .001	0.07	0.01	< .001
PD in Year 1[c]	n/a			0.03	0.01	.015
PD in Year 2[d]	n/a			0.00	0.01	.896
Predictors of the slope[a]						
Grade level	-0.47	0.08	< .001	-0.46	0.08	< .001
Intelligence	-0.01	0.04	.804	-0.02	0.04	.655
Verbal WM	-0.03	0.03	.342	-0.03	0.03	.340
Visual-spatial WM	0.01	0.04	.788	0.02	0.04	.665
PD in Year 1[c]	n/a			-0.06	0.06	.356
PD in Year 2[d]	n/a			0.03	0.07	.640
Correlations[a]						
Intercept with slope	0.09	0.06	.095	0.10	0.05	.059
Intelligence with grade level	0.07	0.02	.002	0.07	0.02	.002
Verbal WM with grade level	0.04	0.03	.090	0.04	0.02	.090
Verbal WM with intelligence	0.39	0.02	< .001	0.40	0.02	< .001
Visual-spatial WM with grade level	0.04	0.02	.087	0.04	0.02	.088
Visual-spatial WM with intelligence	0.37	0.02	< .001	0.37	0.02	< .001
Visual-spatial WM with verbal WM	0.36	0.02	< .001	0.36	0.02	< .001
Intercepts[b]						
Intercept	76.67	0.33	< .001	76.07	0.48	< .001
Slope	7.35	0.20	< .001	7.42	0.36	< .001
Residual variances[b]						
Maths T3	36.42	4.88	< .001	36.07	4.82	< .001
Maths T4	43.05	2.86	< .001	43.29	2.84	< .001
Maths T5	23.86	4.89	< .001	23.44	4.88	< .001
Intercept	103.82	3.71	< .001	102.88	3.66	< .001
Slope	9.53	2.29	< .001	9.66	2.29	< .001
Explained variance						
Maths T3	0.96	0.01	< .001	0.96	0.01	< .001
Maths T4	0.95	0.00	< .001	0.95	0.00	< .001
Maths T5	0.97	0.01	< .001	0.97	0.01	< .001
Intercept	0.88	0.01	< .001	0.88	0.01	< .001
Slope	0.22	0.07	.002	0.22	0.07	.002

Note. WM = working memory, intelligence = nonverbal intelligence. For parsimony, the means (all close to 0 due to centering) and variances of the predictors are omitted from the table.
[a] standardised [b] unstandardised; [c] Cohort 1: long-term effect d Cohort 2: short-term effect

4.3.6 Multiple-group model Year 2

In the Year 2 multiple-group model, two residual variances (mathematics T3 and mathematics T5) were initially negatively estimated in the low-achieving and average-achieving group and were fixed to 0. After this, the multiple-group model had a good fit: χ^2 (49) = 136.392, p < .001, *RMSEA* = .039, *CFI* = 0.996, *TLI* = 0.995, *SRMR* = .035. Again, the results were similar to the overall model, although the predictive value of the covariates and the correlations between them varied somewhat between the achievement groups (see Table 4.8). Similar to the overall model, the multiple group model demonstrated no significant short-term or long-term effect of PD on the slope in any of the achievement groups. Wald tests confirmed that these parameters were similar across achievement groups. Thus, in contrast to our hypothesis, neither long-term nor short-term effects on the mathematics achievement growth could be demonstrated in any of the achievement groups in Year 2.

4.4 Discussion

This large-scale study investigated the effects of a PD programme about differentiation on student achievement growth in mathematics. We hypothesised that the PD programme would have a positive effect on student achievement and that this would be true for students of all achievement levels. Our results provide partial support for these hypotheses: The PD had positive effects on students of all achievement levels in Year 1, but these effects could not be replicated in Year 2.

In Year 1, the overall analysis demonstrated a small but significant positive effect of the PD programme on student achievement growth in mathematics. The multiple-group analysis demonstrated that the direction of effects was positive for all achievement groups, as hypothesised, and that the effect size was similar for low-achieving, average-achieving and high-achieving students. Thus, we found no evidence for differential effects depending upon achievement level. Our findings contrast with some previous studies of naturally occurring ability grouping – without information about differentiation – in which negative effects of being placed in a low-ability group were found (Condron, 2008; Nomi, 2010; reviewed by Deunk et al., 2015).

In line with previous reviews about ability grouping which stressed the importance of adapting instruction to the specific needs of the groups (Kulik & Kulik, 1992; Lou et al., 1996; Slavin, 1987), we believe that an important success factor in our project was that teachers were provided with the skills and knowledge to use ability grouping as a *means* to differentiate instruction rather than as an end in itself. In the PD programme, attention was spent on all four dimensions of adaptive teaching competency, which has been shown to relate positively to student achievement (Vogt & Rogalla, 2009): knowledge about mathematics (e.g., the sequence in which children learn mathematical concepts

Table 4.8 Multiple-group model Year 2

Parameter	Low-achieving *n* = 1029			Average-achieving *n* = 1159			High-achieving *n* = 1285		
	Estimate	*SE*	*p*	Estimate	*SE*	*p*	Estimate	*SE*	*p*
Predictors of the intercept[a]									
Grade level	0.92	0.01	< .001	0.96	0.01	< .001	0.93	0.01	< .001
Intelligence	0.10	0.02	< .001	0.05	0.01	< .001	0.11	0.01	< .001
Verbal WM	0.04	0.01	.002	0.01	0.01	.253	0.04	0.01	.002
Visual-spatial WM	0.05	0.01	< .001	0.03	0.01	.003	0.03	0.01	.012
PD in Year 1[c]	0.04	0.02	.035	0.01	0.01	.475	0.03	0.02	.123
PD in Year 2[d]	0.00	0.02	.864	0.00	0.01	.825	-0.01	0.02	.728
Predictors of the slope[a]									
Grade level	-0.32	0.06	< .001	-0.40	0.05	< .001	-0.24	0.08	.002
Intelligence	0.09	0.05	.050	0.15	0.04	< .001	0.10	0.04	.022
Verbal WM	0.05	0.04	.313	0.08	0.04	.039	0.01	0.05	.847
Visual-spatial WM	-0.01	0.04	.845	-0.02	0.04	.695	0.11	0.05	.023
PD in Year 1 [c]	-0.02	0.07	.819	0.02	0.06	.780	-0.06	0.07	.341
PD in Year 2 [d]	0.02	0.06	.765	0.01	0.06	.887	0.02	0.08	.791
Correlations[a]									
Intercept with slope	0.25	0.05	< .001	0.63	0.03	< .001	0.40	0.08	< .001
Intelligence with grade level	0.01	0.04	.725	0.12	0.04	.002	0.07	0.04	.077
Verbal WM with grade level	0.12	0.04	.002	0.07	0.04	.101	0.03	0.04	.381
Verbal WM with intelligence	0.21	0.03	< .001	0.25	0.03	< .001	0.33	0.03	< .001
Visual-spatial WM with grade level	0.13	0.05	.004	0.08	0.04	.045	0.02	0.03	.528
Visual-spatial WM with intelligence	0.35	0.04	< .001	0.23	0.03	< .001	0.21	0.03	< .001
Visual-spatial WM with verbal WM	0.30	0.04	< .001	0.24	0.03	< .001	0.23	0.03	< .001

Table 4.8 continues on next page

Table 4.8 *Continued*

Parameter	Low-achieving *n* = 1029			Average-achieving *n* = 1159			High-achieving *n* = 1285		
	Estimate	*SE*	*p*	Estimate	*SE*	*p*	Estimate	*SE*	*p*
Intercepts[b]									
Intercept	66.87	0.67	< .001	82.77	0.33	< .001	94.69	0.50	< .001
Slope	8.50	0.51	< .001	7.46	0.37	< .001	5.82	0.41	< .001
Residual variances[b]									
Maths T1	0.00[e]	n/a	n/a	0.00[e]	n/a	n/a	31.00	5.20	< .001
Maths T2	43.05	2.52	< .001	45.46	2.42	< .001	54.88	3.73	< .001
Maths T3	0.00[e]	n/a	n/a	0.00[e]	n/a	n/a	18.78	6.34	.003
Intercept	58.10	3.94	< .001	25.30	1.69	< .001	40.68	2.46	< .001
Slope	21.03	1.93	< .001	16.05	1.34	< .001	15.16	2.38	< .001
Explained variance									
Maths T1	1.00	n/a	n/a	1.00	n/a	n/a	0.94	0.01	< .001
Maths T2	0.92	0.01	< .001	0.90	0.01	< .001	0.88	0.01	< .001
Maths T3	1.00	n/a	n/a	1.00	n/a	n/a	0.96	0.02	< .001
Intercept	0.89	0.01	< .001	0.94	0.01	< .001	0.90	0.01	< .001
Slope	0.11	0.04	.007	0.18	0.04	< .001	0.08	0.04	.047

Note. WM = working memory, intelligence = nonverbal intelligence. For parsimony, the means (all close to 0 due to centering) and variances of the covariates are omitted from the table.

[a] standardised [b] unstandardised [c] Cohort 1: long-term effect [d] Cohort 2: short-term effect [e] fixed to 0

and skills), diagnostic competence (i.e., how to monitor progress and identify educational needs), teaching methods (e.g., how to vary the level of abstraction in response to students' needs) and classroom management (e.g., how to organise within-class ability grouping). In the evaluation questionnaire, teachers indicated that they had learned about all steps in the cycle of differentiation (identification of educational needs, differentiated goals, differentiated instruction, differentiated practice, and evaluation of progress and process; Prast et al., 2015). Moreover, the majority of teachers indicated that they actually used what they had learned in their daily mathematics teaching. We speculate that the positive effects of the PD programme on student achievement can be explained by an increase in teachers' competence for and actual implementation of differentiation, which enabled teachers to better meet their students' educational needs. However, a limitation of this study is that we did not directly investigate the classroom processes underlying the achievement effects since we focused on the final outcome of student achievement. Also, it cannot be determined whether specific components of the intervention were particularly effective. This would require very extensive studies in which specific aspects of the intervention would be systematically varied across multiple experimental conditions. However, due to the interdependence of the steps of the cycle of differentiation, it seems more likely that all aspects of the cycle of differentiation work together than that one isolated component would be effective by itself. In future research, mixed methods could be used to examine in more depth how the PD affects classroom processes and, in turn, student achievement.

In contrast to our hypothesis, the positive effects of the PD in Year 1 could not be replicated in Year 2. One possible explanation is that schools in Cohort 2 were less motivated for the PD programme than schools in Cohort 1 due to the design of the study. When schools registered for the study, most schools were eager to participate in the PD programme. Possibly, schools in Cohort 1, in which the PD programme immediately started, were ready and motivated, whereas schools in Cohort 2 had to wait for one year during which motivation or priorities for PD may have changed. Indeed, one school from Cohort 2 dropped out and several schools in Cohort 3 declined participation in the PD programme when it was offered to them after Year 2. This shows that a school's needs and priorities are dynamic and that a PD programme which suits the needs of a school in one year may not be (as) interesting for the school one or two years later.

Another possible explanation for the smaller effects in Cohort 2 is that teachers of Cohort 2 on average had fewer years of teaching experience at the start of the intervention. Moreover, relatively many teachers were new at the school in the year of the PD. For less experienced teachers and for teachers who just started at a new school, it may be more challenging to implement differentiation since they may need to spend attention first on more basic issues such as classroom management and (new) everyday routines. In addition, relatively many schools in Cohort 2 started to use another mathematics curriculum during the course of the study. This may have drawn teachers' attention towards implementation of the new curriculum rather than to the implementation of differentiation (although school administrators themselves generally viewed it as an asset that these could be combined). These explanations illustrate that this study was situated in the dynamic context of daily practice in schools. This is both a strength and a limitation: while it promotes the practical validity of the findings, it diminishes the experimental control.

4.4.1 Implications and future research

This study was designed to have strong links to educational practice. Therefore, the PD programme was designed in collaboration with experienced teacher trainers who could bridge theory and practice. Moreover, this was the first large-scale study to investigate achievement effects of a PD programme about differentiation in mathematics. This question has large practical relevance because, although differentiation and PD about this topic are often promoted by policy makers, little was known about the effects of such interventions. The results show that PD about differentiation can improve student achievement, but that such achievement effects are not guaranteed.

Probably, much depends on whether teachers are able to apply what they learned during the PD in daily practice. We noticed during the PD programme that, while most teachers already implemented some aspects of differentiation such as tiered tasks if those were provided by the mathematics curriculum, the challenge of the PD programme was to increase the *quality* of differentiation by (1) implementing differentiation more systematically, using the full cycle of differentiation for students of all achievement levels and (2) improving the match between diagnosed educational needs and instructional adaptations. This required substantial mathematical knowledge, for example regarding the typical sequence of learning mathematical concepts and operations (enabling teachers to move back to more fundamental steps if necessary). While in-service teacher education may be a way to develop such knowledge, pre-service teacher education could also strive to equip teachers with more systematic knowledge about mathematics and didactics of mathematics before they enter the workplace. PD for in-service teachers could then focus on more advanced components of adaptive teaching competency for which this knowledge is required, such as how to use refined diagnostics to find the most appropriate instructional adaptations for a particular student. Besides adaptive teaching competency, the implementation of differentiation is also influenced by contextual factors such as the availability of appropriate instructional materials and preparation time (Roiha, 2014). School administrators could support their teachers in the implementation of differentiation by facilitating such practical aspects (Puzio, Newcomer, & Goff, 2015).

A limitation of the current study is that we did not examine directly how teacher and student learning processes influenced students' learning outcomes. While previous case studies have reported about the process of starting to implement differentiation (e.g., Brimijoin, 2002), future research could examine the effects of PD on adaptive teaching competency, implementation of differentiation, student learning processes and student achievement jointly and investigate how these effects interact over time. Small-scale studies using both quantitative and qualitative measures may be suitable to unravel such processes.

Another issue is whether the achievement effects in the current study were practically significant. The effect sizes were quite small but if this modestly higher achievement growth could be sustained over multiple years, the cumulative effect would be substantial. However, the higher achievement growth was not sustained in Cohort 1 after the PD programme had ended. This may require prolonged PD (c.f. VanTassel-Baska et al., 2008). To increase the effect sizes, future research could also investigate how technological applications for differentiation such as Accelerated Math (Ysseldyke et al., 2003) and PD about differentiation could be combined. Technological applications could be used to support and relieve teachers wherever possible, complemented with PD to develop teachers' competencies in qualitative analysis and refined instructional adaptations.

In conclusion, the results of this study show that PD about differentiation in mathematics has the potential to raise the achievement of all students. This is consistent with educational theories including the zone of proximal development (Vygotsky, 1978), aptitude-treatment interaction (Cronbach & Snow, 1977), and adaptive teaching (Corno, 2008) which propose that educational needs vary based on achievement level and that adapting education to those diverse needs leads to more effective learning. Our results indicate that schoolwide PD about systematic implementation of differentiation using the cycle of differentiation may have positive effects over and above the spontaneous adaptations that many teachers already make by themselves. Despite the drawbacks discussed above, we think that these results are sufficiently promising to continue this line of research.

References

Alloway, T., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology*, *28*, 725-734. doi:10.1080/01443410802243828

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (pp. 548-556). Oxford, UK: Elsevier.

Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal*, *44*, 701-731. doi:10.3102/0002831207306743

Brimijoin, K. (2002). *Expertise in differentiation: A preservice and inservice teacher make their way*. Charlottesville, VA: University of Virginia.

Brown, J., & Morris, D. (2005). Meeting the needs of low spellers in a second-grade classroom. *Reading & Writing Quarterly*, *21*, 165-184. doi:10.1080/10573560590915969

Condron, D. J. (2008). An early start: Skill grouping and unequal reading gains in the elementary years. *Sociological Quarterly*, *49*, 363-394. doi:10.1111/j.1533-8525.2008.00119.x

Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, *43*, 161-173. doi:10.1080/00461520802178466

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper-Perennial.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13-21. doi:10.1016/ j.intell.2006.02.001

Deunk, M., Doolaard, S., Smale-Jacobse, A., & Bosker, R. J. (2015). *Differentiation within and across classrooms: A systematic review of studies into the cognitive effects of differentiation practices*. Groningen, The Netherlands: GION.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Friso-Van den Bos, I., Van der Ven, S. H. G., Kroesbergen, E. H., & Van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, *10*, 29-44. doi:10.1016/j.edurev.2013.05.003

Fullan, M. (2002). The change leader. *Educational Leadership*, *59*(8), 16-21.

Gal'perin, P. J. (1969). Stages in the development of mental acts. In M. Cole, & I. Maltzman (Eds.), *A handbook of contemporary Soviet psychology* (pp. 249-273). New York, NY: Basic Books.

Gamoran, A. (1992). Is ability grouping equitable: Synthesis of research. *Educational Leadership*, *50*, 11-17.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*, 1539-1552. doi:10.1037/a0025510

Grimes, K. J., & Stevens, D. D. (2009). Glass, bug, mud. *Phi Delta Kappan*, *90*, 677-680. doi:10.1177/003172170909000914

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55. doi:10.1080/10705519909540118

Inspectie van het Onderwijs. (2015). *Beginnende leraren kijken terug* [Beginning teachers look back]. Utrecht, The Netherlands: Inspectie van het Onderwijs.

Janssen, J., Scheltens, F., & Kraemer, J. M. (2005a). *Rekenen-wiskunde groep 3–8: Handleidingen* [Mathematics test grade 1 through 6: Manuals]. Arnhem, The Netherlands: Cito.

Janssen, J., Scheltens, F., & Kraemer, J. M. (2005b). *Leerling- en onderwijsvolgsysteem rekenen-wiskunde* [Student progress monitoring system mathematics]. Arnhem, The Netherlands: Cito.

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS rekenenwiskunde voor groep 3 tot en met 8* [Scientific justification of the mathematics test for grade 1 through grade 6.]. Arnhem, The Netherlands: Cito.

Johnsen, S. K., Haensly, P. A., Ryser, G. A. R., & Ford, R. F. (2002). Changing general education classroom practices to adapt for gifted students. *Gifted Child Quarterly*, *46*, 45-63. doi: 10.1177/001698620204600105

Keuning, J., Van Boxtel, H., Lansink, N., Visser, J., Weekers, A., & Engelen, R. (2015). *Actualiteit en kwaliteit van normen: Een werkwijze voor het normeren van een leerlingvolgsysteem* [Up-to-dateness and quality of norms: A method to develop norms for a student progress monitoring system.]. Arnhem, The Netherlands: Cito.

Klingner, J. K., Ahwee, S., Pilonieta, P., & Menendez, R. (2003). Barriers and facilitators in scaling up research-based practices. *Exceptional Children*, *69*, 411-429. doi:10.1177/001440290306900402

KNAW. (2009). *Rekenonderwijs op de basisschool: Analyse en sleutels tot verbetering* [Mathematics education in primary school: Analysis and keys to improvement]. Amsterdam, The Netherlands: KNAW.

Koellner, K., & Jacobs, J. (2015). Distinguishing models of professional development: The case of an adaptive model's impact on teachers' knowledge, instruction, and student achievement. *Journal of Teacher Education*, *66*, 51-67. doi:10.1177/0022487114549599

Kulik, J. A., & Kulik, C. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly*, *36*, 73-77. doi:10.1177/001698629203600204

Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching & Teacher Education*, *25*, 12-23. doi:10.1016/j.tate.2008.06.001

Leonard, J. (2001). How group composition influenced the achievement of sixth-grade mathematics students. *Mathematical Thinking and Learning*, *3*, 175-200. doi:10.1080/10986065.2001.9679972

Lou, Y., Abrami, P. C., & Spence, J. C. (2000). Effects of within-class grouping on student achievement: An exploratory model. *Journal of Educational Research*, *94*, 101-112. doi:10.1080/00220670009598748

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & D'Appolonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*, 423-458. doi:10.3102/00346543066004423

McDonald Connor, C., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science*, *315*, 464-465. doi:10.1126/science.1134513

McDonald Connor, C., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P. S., ... Schatschneider, C. (2011a). Testing the impact of child characteristics x instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly*, *46*, 189-221. doi:10.1598/RRQ.46.3.1

McDonald Connor, C., Morrison, F. J., Schatschneider, C., Toste, J. R., Lundblom, E., Crowe, E. C., & Fishman, B. (2011b). Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness*, *4*, 173-207. doi:10.1080/19345747.2010.510179

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*, 114-140. doi: 10.1037/met0000078

Murata, A. (2011). Introduction: Conceptual overview of Lesson Study. In L. C. Hart, A. Alston, & A. Murata (Eds.), *Lesson Study research and practice in mathematics education* (pp. 1-12). London, UK: Springer.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nomi, T. (2010). The effects of within-class ability grouping on academic achievement in early elementary years. *Journal of Research on Educational Effectiveness*, *3*, 56-92. doi:10.1080/19345740903277601

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*, 237-257. doi:10.3102/01623737026003237

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2015). Differentiation in primary school mathematics: Expert recommendations and teacher self-assessment. *Frontline Learning Research*, *3*, 90-116. doi:10.14786/flr.v3i2.163

Puzio, K., Newcomer, S. N., & Goff, P. (2015). Supporting literacy differentiation: The principal's role in a community of practice. *Literacy Research and Instruction*, *54*, 135-162. doi: 10.1080/19388071.2014.997944

Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's standard progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press.

Rogers, K. B. (2007). Lessons learned about educating the gifted and talented: A synthesis of the research on educational practice. *Gifted Child Quarterly*, *51*, 382-396. doi:10.1177/0016986207306324

Roiha, A. L. (2014). Teachers' views on differentiation in content and language integrated learning (CLIL): Perceptions, practices and challenges. *Language and Education*, *28*, 1-18. doi: 10.1080/09500782.2012.748061

Roy, A., Guay, F., & Valois, P. (2013). Teaching to address diverse learning needs: Development and validation of a differentiated instruction scale. *International Journal of Inclusive Education*, *17*, 1186-1204. doi:10.1080/13603116.2012.743604

Schram, E., Van der Meer, F., & Van Os, S. (2013). *Omgaan met verschillen: (G)een kwestie van maatwerk* [Responding to differences: (Not) a matter of customisation]. Enschede, The Netherlands: SLO.

Schweizer, K., Goldhammer, F., Rauch, W., & Moosbrugger, H. (2007). On the validity of Raven's matrices test: Does spatial ability contribute to performance? *Personality and Individual Differences*, *43*, 1998-2010. doi:10.1016/j.paid.2007.06.008

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. *Review of Educational Research*, *57*, 293-336. doi:10.3102/00346543057003293

Smeets, E., Ledoux, G., Regtvoort, A., Felix, C., & Mol Lous, A. (2015). *Passende competenties voor passend onderwijs: Onderzoek naar competenties in het basisonderwijs* [Adaptive competencies for inclusive education: Research about competencies in primary school]. Nijmegen, The Netherlands: ITS.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary* (3rd ed.). New York, NY: Oxford University Press.

Tieso, C. L. (2003). Ability grouping is not just tracking anymore. *Roeper Review*, *26*, 29-36. doi: 10.1080/02783190309554236

Tieso, C. L. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted*, *29*, 60-89.

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Jolani, S., & Van Luit, J. E. H. (2016). The monkey game: A computerized verbal working memory task for self-reliant administration in primary school children. *Behavior Research Methods*, *48*, 756-771. doi:10.3758/s13428-015-0607-y

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Prast, E. J., & Van Luit, J. E. H. (2015). Validity and reliability of an online working memory task for self-reliant administration in school-aged children. *Behavior Research Methods*, *47*, 708-719. doi:10.3758/s13428-014-0469-8

Van Groenestijn, M., Borghouts, C., & Janssen, C. (2011). *Protocol ernstige rekenwiskunde-problemen en dyscalculie* [Protocol severe mathematics difficulties and dyscalculia]. Assen, The Netherlands: Van Gorcum.

VanTassel-Baska, J., Feng, A. X., Brown, E., Bracken, B., Stambaugh, T., French, H., ... Bai, W. (2008). A study of differentiated instructional change over 3 years. *Gifted Child Quarterly*, *52*, 297-312. doi:10.1177/0016986208321809

VanTassel-Baska, J., & Stambaugh, T. (2005). Challenges and possibilities for serving gifted learners in the regular classroom. *Theory into Practice*, *44*, 211-217. doi:10.1207/s15430421tip4403_5

Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, *25*, 1051-1060. doi:10.1016/j.tate.2009.04.002

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Ysseldyke, J., & Bolt, D. M. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. *School Psychology Review*, *36*, 453-467.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk*, *8*, 247-265. doi:10.1207/S15327671ESPR0802_4

5

Relations between mathematics achievement and motivation in students of diverse achievement levels

Emilie J. Prast
Eva Van de Weijer-Bergsma
Milica Miočević
Evelyn H. Kroesbergen
Johannes E. H. Van Luit

Manuscript submitted for publication

Author contributions: All authors except MM designed research; EvdW and EP collected the data under supervision of HvL and EK; EP and MM analysed the data; EP wrote the paper; EvdW, MM, EK and HvL provided feedback on the paper.

Abstract

Motivation and achievement are known to be related, but the nature of this relation is complex. This study investigated the relations between achievement and several core aspects of motivation for mathematics in primary school: self-efficacy, self-concept, task value, and mathematics anxiety (N = 4306 students of grade 2 – 6). Moreover, it was investigated whether these relations were similar or different for low-achieving, average-achieving, and high-achieving students. Students completed a standardised mathematics achievement test at T1 and T3 and a mathematics motivation questionnaire at T2. Working memory was measured as a covariate. Self-efficacy and self-concept were combined into a single perceived competence variable due to their high intercorrelation. T1 achievement positively predicted perceived competence and task value and negatively predicted mathematics anxiety. Only perceived competence had a significant effect on T3 achievement after controlling for T1 achievement and working memory, and significantly mediated between previous and subsequent achievement. This pattern of effects was largely similar across achievement groups, although the effects of previous achievement on task value and perceived competence were stronger for high-achieving students. In each achievement group, perceived competence was the only motivational variable with a significant effect on subsequent achievement over and above the effects of previous achievement.

5.1 Introduction

Motivation and achievement are closely related: Students tend to feel more competent in the school subjects in which they achieve well and value these subjects more highly too (Denissen, Zarrett, & Eccles, 2007). However, the relations between motivation and achievement seem to be complex. In many cases, the relations between motivation and achievement are theorised to be reciprocal. Thus, motivation is not only influenced by previous achievement but also predicts subsequent achievement (e.g., Marsh & Martin, 2011). Moreover, these relations seem to differ depending on the aspect of motivation under study. For example, the effect of self-efficacy on subsequent achievement has been demonstrated repeatedly (e.g., Valentine, DuBois, & Cooper, 2004) whereas the effect of task value on subsequent achievement has been called into question (e.g., Garon-Carrier et al., 2016). Another factor adding to the complexity is that motivational characteristics may interact with other individual learner characteristics such as working memory (e.g., Ramirez, Chang, Maloney, Levine, & Beilock, 2016). An individual learner characteristic that receives particular attention in the current study is achievement level. Since the educational experiences of low-achieving, average-achieving, and high-achieving students differ substantially, the relations between motivation and achievement might also differ depending on the achievement level of the student. More knowledge about these complex relations might provide directions for differentiating instruction based on students' motivational needs.

In the current study, we examined whether several core aspects of motivation for mathematics – self-efficacy, self-concept, task value, and mathematics anxiety – were predicted by previous achievement in mathematics. Moreover, we examined how these motivational variables jointly predicted subsequent mathematics achievement and whether they mediated between previous and subsequent mathematics achievement. Finally, we examined whether these relations between motivation and achievement were similar or different for low-achieving, average-achieving, and high-achieving students within heterogeneous primary schools.

5.1.1 Perceived competence: Self-efficacy and self-concept

Students' perceptions about their own competence are an important component of many motivational theories, although several different terms with slightly different meanings are used (e.g., self-efficacy, self-concept, expectancy for success, perceived competence). In the current study, we focus on self-efficacy (Bandura, 1977, 1997) and self-concept (Shavelson, Hubner, & Stanton, 1976) and, in line with Hughes, Galbraith and White (2011), we use the term perceived competence to refer to their common core. Despite this common core, some subtle conceptual differences have been established (Bong & Skaalvik, 2003):

Self-efficacy refers to a person's self-estimated capacity to perform a certain task relative to an absolute performance criterion ('I can…')), whereas self-concept comprises a more affective evaluation of a person's own capacities relative to a normative standard ('I am good at…'). Both self-efficacy and self-concept are moderately to strongly correlated to academic achievement, with larger correlations if the motivational and achievement variables are measured in the same domain (e.g., mathematics; Huang, 2012).

Theoretically, the relation between perceived competence and achievement is supposed to be reciprocal: Previous achievement is theorised to be an important source of self-efficacy (Bandura, 1997) and self-concept (Marsh & Martin, 2011). In turn, high self-efficacy and self-concept are supposed to promote adaptive learning behaviours such as persistence, which should have a positive effect on future achievement (Marsh & Martin, 2011; Wigfield & Eccles, 2000). Thus, self-efficacy and self-concept are supposed to mediate the relation between previous and subsequent achievement, with positive effects of self-efficacy and self-concept on subsequent achievement after controlling for previous achievement.

For self-concept, several empirical studies have tested a reciprocal effects model (described by Marsh & Martin, 2011), with somewhat mixed findings. In general, the effects of achievement on self-concept seem to be larger than vice versa (e.g., Chen, Yeh, Hwang, & Lin, 2013; Huang, 2011; Möller et al., 2014; Preckel, Niepel, Schneider, & Brunner, 2013). Regarding the effects of self-concept on achievement, one meta-analysis found a small but significant positive effect on achievement after controlling for previous achievement (Valentine et al., 2004), but this effect was not significant in a later meta-analysis (Huang, 2011). Findings of more recent single studies are also mixed: Some studies did find small but significant effects of self-concept on achievement after controlling for previous achievement (Chen et al., 2013; Kriegbaum et al., 2015; Niepel, Brunner, & Preckel, 2014), but other longitudinal[1] studies found no or only partial evidence for such a relation (Möller et al., 2014; Preckel et al., 2013; Viljaranta et al., 2014).

For self-efficacy, the available studies with control for previous achievement yielded more consistent results: Several studies have reported small to moderate significant effects from self-efficacy on achievement after controlling for previous achievement (Fast et al., 2010; Jungert, Hesser, & Träff, 2014; Kriegbaum et al., 2015; Valentine et al., 2004). Moreover, self-efficacy was shown to mediate the relation between previous and subsequent achievement in typically achieving students (Jungert et al., 2014).

[1] Throughout this article, we use the term longitudinal to refer to studies with at least two measurement occasions (enabling statistical control for previous achievement), to distinguish these from cross-sectional studies in which motivation and achievement data were collected at a single timepoint.

5.1.2 Task value

Expectancy-value theories of achievement motivation posit that motivation for a task is not only determined by the student's expectancy for success based on perceived competence but also by the degree to which the student values the task (Wigfield & Eccles, 2000). Three major components of task value are intrinsic or interest value (enjoyment gained from engaging in the task), personal or attainment value (perceived importance of the task), and utility value (perceived usefulness of the task) (Wigfield & Cambria, 2010). Theoretically, task value is supposed to promote adaptive learning behaviours such as persistence and enhance performance (Wigfield & Eccles, 2000).

However, recent empirical studies have suggested that the effects of task value[2] on subsequent achievement may be rather limited after controlling for previous achievement: three longitudinal studies found no significant effects (Garon-Carrier et al., 2016; Jõgi, Kikas, Lerkkanen, & Mägi, 2015; Viljaranta, Tolvanen, Aunola, & Nurmi, 2014), whereas two other studies found significant but very small effects of task value on subsequent achievement (Corpus, McClintic-Gilbert, & Hayenga, 2009; Kriegbaum, Jansen, & Spinath, 2015). Task value does seem to be related to previous achievement: Several longitudinal studies have reported small to moderate positive effects from previous achievement on subsequent task value (Corpus et al., 2009; Garon-Carrier et al., 2016; Gniewosz, Eccles, & Noack, 2015; Jõgi et al., 2015; Von Maurice, Dörfler, & Artelt, 2014; by exception, Viljaranta et al. (2014) found no significant effect). Thus, students who perform well at a task are likely to value it more in the future but recent studies suggest that task value does not necessarily enhance future achievement.

5.1.3 Mathematics anxiety

Mathematics anxiety refers to feelings of worry, fear and tension which arise when engaging in mathematical activities (Suinn & Winston, 2003). Although the construct of mathematics anxiety is more affective than motivational, it was also included in the present study because it is theoretically and empirically related to perceived competence, task value, and achievement (Bong, Cho, Ahn, & Kim, 2012); Krinzinger, Kaufmann, & Willmes, 2009; Van der Beek, Van der Ven, Kroesbergen, & Leseman, 2017).[3] In theory, mathematics anxiety is reciprocally related to mathematics achievement: Frequent failure to understand or perform mathematics tasks is supposed to provoke mathematics anxiety, which is in turn theorised to have a negative influence on mathematics performance because the anxiety

[2] In this article, we use the term 'task value', but we also refer to studies in which the strongly related constructs of interest and intrinsic motivation for mathematics were investigated using similar measures; see Wigfield & Cambria (2010) for a review of similarities and differences between these constructs

[3] For the sake of readability, we refer to mathematics anxiety as one of the motivational variables in our study (although it would be more correctly classified as an affective variable).

is associated with worrisome thoughts (taking away attention from the calculation process) and avoidance behaviour (reducing the amount of practice; Krinzinger et al., 2009).

Research about the relation between mathematics anxiety and achievement is mostly cross-sectional in nature. Two meta-analyses reported moderate negative correlations between mathematics anxiety and mathematics achievement in elementary and secondary school students (Hembree, 1990; Ma, 1999). More recent cross-sectional studies found small to moderate negative effects from mathematics anxiety on achievement (Wang et al., 2015; Wu, Barth, Amin, Malcarne, & Menon, 2012) and strong negative effects from achievement on mathematics anxiety (Birgin, Baloğlu, Çatlıoğlu, & Gürbüz, 2010). The relation between mathematics anxiety and achievement seems to be moderated by the working memory (WM) load of the strategies used to solve the task, with stronger relations for tasks and strategies requiring high WM capacity (Ramirez et al., 2016; Ramirez, Gunderson, Levine, & Beilock, 2013; Wu et al., 2012). A possible explanation is that mathematics anxiety places an additional burden on WM, thus interfering more with strategies which already tax WM (Ashcraft, 2002; Ramirez et al., 2016). Remarkably, the negative effects of mathematics anxiety seem to be stronger for students with *high* WM capacities, probably due to their tendency to use WM-intensive strategies (Ramirez et al., 2016; Ramirez et al., 2013).

Longitudinal studies about the relation between mathematics anxiety and achievement are very scarce, but support the idea that mathematics anxiety predicts future achievement only for tasks with a high WM load and for students with above-average WM capacity. Specifically, two studies in the early primary grades found no effect from previous mathematics anxiety on basic addition and subtraction tasks (Krinzinger et al., 2009; Vukovic, Kieffer, Bailey, & Harari, 2013). Vukovic et al. (2013) did find a moderate negative effect on more complex application tasks, but only for students with above-average working memory skills. Given these indications for a potential interaction between mathematics anxiety and working memory, we included working memory in our study not only as a covariate but also to test this potential interaction with mathematics anxiety.

5.1.4 Interrelations and relative predictive power of self-efficacy, self-concept, task value, and mathematics anxiety

Previous studies have shown that task value is positively related to perceived competence (Bong et al., 2012; Denissen et al., 2007), whereas mathematics anxiety is negatively related to perceived competence and task value (Bong et al., 2012; Krinzinger et al., 2009; Van der Beek et al., 2017). When multiple aspects of motivation are modelled as predictors of achievement, perceived competence generally has larger predictive power than task value and mathematics anxiety (Bong et al., 2012; Kriegbaum et al., 2015; Spinath, Spinath, Harlaar, & Plomin, 2006). When both self-efficacy and self-concept are simultaneously included as separate aspects of perceived competence, the two constructs are typically strongly related - sometimes so strongly that multicollinearity problems arise (Bong et al., 2012; Marsh, Dowson, Pietsch, & Walker, 2004). This has given rise to the question whether the two constructs are empirically distinguishable (Hughes et al., 2011). Nevertheless, recent studies provided indications that self-efficacy and self-concept do explain some unique variance in achievement (Huang, 2012; Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014). Previous results did not consistently indicate whether self-efficacy or self-concept is a stronger predictor of achievement (see the meta-analyses by Huang, 2012 and Valentine et al., 2004; as well as more recent findings by Parker et al., 2014). Therefore, both self-efficacy and self-concept were included in the present study.

Few studies have examined how several motivational aspects together longitudinally predict achievement (with control for previous achievement). In one large-scale longitudinal study with secondary school students (Kriegbaum et al., 2015), self-efficacy, self-concept, and multiple components of task value (interest, enjoyment, and utility value) were first separately modelled as predictors of achievement and each of these variables was found to predict achievement after controlling for previous achievement and intelligence. A subsequent relative weights analysis demonstrated that the effects were larger for self-efficacy and self-concept than for task value. For self-efficacy, the predictive power depended upon the specificity of the questions: Task-specific self-efficacy (measured with questions including examples of mathematics tasks) was a stronger predictor of mathematics achievement than self-concept for mathematics (in general), whereas self-efficacy for mathematics (in general) had less predictive power.

5.1.5 Academic achievement level as a moderator

It is conceivable that the relations between motivation and achievement are different for students of different academic achievement levels because the achievement experiences of low-achieving and high-achieving students within heterogeneous classrooms differ substantially: Low-achieving students may be expected to experience failure relatively often whereas high-achieving students may be used to experiencing success. For example, Hampton and Mason (2003) found that students with learning disabilities experienced less (previous) accomplishment, less positive reinforcement from others, fewer role models, and more anxiety than typically-achieving students. The relations between motivation and achievement might be moderated by achievement level in several ways. First, the effects of previous achievement on motivation might depend on achievement level. For example, the importance of achievement on a standardised test as a source of information in the formation of motivation towards mathematics might be different for low-achieving versus high-achieving students. Second, motivation might be generally more important for one group of students. For example, high motivation might be especially beneficial for subsequent achievement in low-achieving students, who might need more persistence to reach an adequate achievement level. Third, the relative importance of various aspects of motivation might be different depending upon achievement level. For example, self-efficacy might be relatively more important for one achievement group whereas task value might be relatively more important for another achievement group. Knowledge about these potential differences between achievement groups could have implications for educational practice. For example, if the relative importance of various aspects of motivation would indeed differ between achievement groups this might provide indications for differentiated instruction: Teachers might attempt to foster those aspects of motivation which are most strongly related to subsequent achievement in a particular achievement group.

However, only few studies examined academic achievement level as a potential moderator of the relations between motivation and achievement. For self-concept, one study (Möller et al., 2014) compared students in the academic and vocational track of secondary school. The effect of achievement on self-concept was smaller in the vocational track, but only when grades (rather than standardised achievement tests) were used as the achievement indicator. The effects of self-concept on achievement were similar across groups (positive in direction but nonsignificant). For self-efficacy, one study (Jungert et al., 2014) found that whereas mathematics achievement and self-efficacy were reciprocally related in typically achieving students, these relations were not significant for low-achieving students. For task value, we did not find previous studies comparing students of diverse achievement levels but one study did compare students of low versus typical general ability levels (Jõgi et al., 2015). Previous mathematics achievement was a significantly stronger predictor of task value for low-ability students. In both groups, there was no significant effect from task value on subsequent achievement. For mathematics anxiety, Krinzinger et al. (2009) found no indications that the development of mathematics anxiety and achievement over time was different for students of different anxiety or achievement levels – in other words, the lack of effects of mathematics anxiety on achievement for the total sample was replicated in subsamples of students with high anxiety or low mathematics performance. Finally, a study comparing students of three achievement levels in Dutch secondary schools did find mean differences in self-concept, enjoyment, and mathematics anxiety (high-achieving students scored more favourably on all aspects) but the relations between these constructs and achievement were similar for students of diverse achievement levels (Van der Beek et al., 2017).

5.1.6 Research questions and hypotheses

Previous studies have already provided many insights into the relations of self-efficacy, self-concept, task value, and mathematics anxiety with achievement, despite some inconsistencies in the results. However, few studies have investigated all of these aspects of motivation together. This would be relevant, since previous studies have provided indications that the relations between various aspects of motivation and achievement are also dependent on the other motivational aspects included in the model. For example, Spinath et al. (2006) found that the effect of task value on achievement was no longer significant when self-concept was also included in the model. Studies that did investigate multiple aspects of motivation often used cross-sectional rather than longitudinal designs (e.g., Bong et al., 2012; Spinath et al., 2006) or focused on secondary school students (Kriegbaum et al., 2015). Thus, there is still a need for studies which investigate the longitudinal relations between these various aspects of motivation and achievement in primary school. Moreover, little is known about whether these relations are similar or different for students of diverse achievement levels. The few available studies were mostly based on secondary school students (Möller et al., 2014; Van der Beek et al., 2017) or small samples (Jungert et al., 2014; Krinzinger et al., 2009) and differentiated only between low-achieving vs. typically-achieving students (except Van der Beek et al., 2017). However, these relations might be different within heterogeneous primary schools and for high-achieving (vs. typically-achieving or low-achieving) students. Therefore, there is still a need for studies that differentiate between students of multiple achievement levels (e.g., low-achieving, average-achieving, high-achieving) within heterogeneous primary schools. Due to the broad range of achievement levels within heterogeneous primary school classrooms (compared to tracked secondary school classrooms), the motivational differences between these students may also be larger.

In the current study, we examined whether mathematics performance on a standardised achievement test predicted subsequent self-efficacy, self-concept, task value, and mathematics anxiety. Moreover, we examined how self-efficacy, self-concept, task value, and mathematics anxiety jointly predicted subsequent mathematics achievement and whether the motivational variables mediated between previous and subsequent mathematics achievement. In addition, we examined whether the effect of mathematics anxiety on achievement was moderated by working memory level. Finally, we explored whether these relations were similar or different for low-achieving, average-achieving, and high-achieving students.

Based on the literature discussed in this introduction, we hypothesised that:

- Previous mathematics achievement would positively predict self-efficacy, self-concept, and task value and negatively predict mathematics anxiety
- Self-efficacy and self-concept would also predict subsequent achievement, thus mediating between previous and subsequent achievement
- Task value would not predict subsequent achievement
- Mathematics anxiety would predict subsequent achievement negatively, but perhaps only for students with high WM capacity

Given the scarce literature about achievement level as a potential moderator of the relations between motivation and achievement, the investigation of this moderator was exploratory and not guided by specific hypotheses.

5.2 Method

5.2.1 Participants

Data were collected in the context of the large-scale project GROW (see also Prast, Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2015). Thirty-two schools spread across the Netherlands volunteered to participate in this project about differentiation in primary mathematics education[4]. The schools were diverse in terms of school size (*M* = 209 students per school, range 52 to 550) and mathematics curriculum used. The sample consisted of 4306 students (50.7% male) nested in 184 classes (mean class size = 23 students) from grade 2 through 6. All grade levels were equally represented with about 20% of the sample in each grade level. Mean age at the beginning of the study was 9.45 years (*SD* = 1.53).

5.2.2 Measures

5.2.2.1 Mathematics achievement

Mathematics achievement was measured using the Cito mathematics tests (Janssen, Scheltens, & Kraemer, 2005). These are national Dutch tests which are commonly administered at the middle and end of each schoolyear to monitor students' progress in mathematics throughout primary school. For each grade level, different versions with developmentally appropriate tasks for both the middle and end of the schoolyear have been developed. In all versions, five main domains are covered: (a) numbers and number relations, covering the structure of the number line and relations between numbers, (b) addition and subtraction, (c) multiplication and division, (d) complex mathematics applications, often involving multiple mathematical manipulations, and (e) measuring (e.g., weight and length). From mid grade 2 to mid grade 6, the following domains are added successively: (f) estimation, (g) time, (h) money, (i) proportions, (j) fractions, and (k) percentages. The reliability coefficients of the different versions range from .91 to .97 (Janssen, Verhelst, Engelen, & Scheltens, 2010). Based on the means and standard deviations for each grade-level test in a nationally representative sample (Keuning et al.,

[4] In ten of the schools of the current sample, teachers participated in a professional development programme about differentiated mathematics education (see also Prast, Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2018). The remaining 22 schools were in a control condition. We checked whether the results of the current study were similar for students in experimental and control schools. This was the case, so experimental schools were retained in the sample.

2015), the scores on each grade-level test were converted into z-scores (0 = an average score compared to the national norms for students in that grade).

5.2.2.2 Motivation

Motivation for mathematics was assessed with the Mathematics Motivation Questionnaire for Children (MMQC; Prast, Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2012). This self-report questionnaire was designed to measure several aspects of motivation for mathematics in primary school students and includes 24 items about self-efficacy (6 items), self-concept (6 items), task value (7 items), and mathematics anxiety (5 items). All items are rated on the following four-point scale: 1 = NO! (strongly disagree), 2 = no (disagree), 3 = yes (agree), 4 = YES! (strongly agree). Subscale scores are computed by averaging the scores on all items belonging to each subscale (self-efficacy, self-concept, task value and mathematics anxiety). The self-efficacy items concern students' perceived ability to perform mathematics-related tasks. Since the questionnaire was designed for a broad range of grade levels, the questions do not refer to specific mathematical content but to mathematics tasks in general, e.g., "When the teacher explains the first sum, can you do the next sums without help?". In the self-concept items, students are asked to evaluate their own competence in mathematics, e.g., "Are you good at mathematics?". For task value, most items measure the intrinsic or interest value component of task value (e.g., "Do you enjoy doing mathematics?"), but the questionnaire also includes one item each about utility value ("Does it seem handy to you to be good at mathematics?") and personal value ("Do you find mathematics important?"). The mathematics anxiety items concern anxious thoughts and feelings during the mathematics lesson, e.g., "Are you afraid to make mistakes during the mathematics lesson?".

A confirmatory factor analysis had a good fit: *RMSEA* = 0.057, *CFI* = 0.978, *TLI* = 0.975. The chi-square test was significant (χ^2 (246) = 3568.05, $p < .001$), which was expected given the large sample size. All items loaded on their designated factors (range 0.68 – 0.95). Only the factor loadings of the two task value items about personal value (0.44) and utility value (0.49) were somewhat low, but these items were retained since they did represent meaningful aspects of task value in addition to the other items which assessed interest value. Moreover, the internal consistency of the subscale including all task value items was good: α = .87. The other subscales also had a good internal consistency (self-efficacy: α = .81; self-concept: α = .91; mathematics anxiety: α = .84). We investigated the test-retest reliability of the MMQC in a sample of 75 students with a one-week interval, with good results: reliability coefficients ranged from $r = .82$ for self-concept to $r = .93$ for task value.

5.2.2.3 Working memory

Working memory was measured to be included as a covariate and to test a potential interaction between WM and mathematics anxiety. Working memory is an important predictor of mathematics achievement (Friso-Van den Bos, Van der Ven, Kroesbergen, & Van Luit, 2013), with stronger effects on subsequent achievement than general intelligence (Alloway & Alloway, 2010). The Lion game is a visual–spatial complex span task suitable for self-reliant online administration (Van de Weijer-Bergsma, Kroesbergen, Prast, & Van Luit, 2015). Students are presented with a 4 × 4 matrix on the computer screen. In each trial, eight lions of different colours are consecutively presented at different locations in the matrix. Students should remember the last location where a lion of a certain colour has appeared. The Lion game consists of five levels in which working memory load is manipulated by increasing the number of lions (one through five) that students should remember. A mean proportion correct score indicating the proportion of lions recalled in the correct serial position was calculated. To control for the linear and quadratic effects of age, ageresidualised scores were created by regressing the proportion correct score on age and age-squared and saving the unstandardised residuals. The Lion game has demonstrated very good internal consistency ($\alpha = .90$; Van de Weijer-Bergsma et al., 2015). In addition, the Lion game has been shown to correlate ($r = .51 – .59$, $p < .001$) with the individually administered Automated Working Memory Assessment (Alloway, Gathercole, Kirkwood, & Elliott, 2008) and to predict subsequent mathematics achievement ($\beta = .15$, $p < .001$; Van de Weijer-Bergsma et al., 2015).

5.2.3 Procedure

The mathematics tests were administered by the classroom teacher as part of the standard national achievement testing procedure in June 2012 (T1) and February 2013 (T3). The motivation questionnaire was group-administered in the classroom under supervision of a research assistant in September 2012 (T2). In grades 2 and 3, the research assistant read each question aloud, after which the students wrote down their answer. In grades 4 through 6, students completed the questionnaire independently after receiving instructions. The Lion game was administered online at T2: Teachers were asked to assign the task and monitor that all students completed the task self-reliantly. `

5.2.4 Data analysis

Data were analysed in M*plus* (version 7.3, Muthén & Muthén, 1998–2012) using structural equation modelling. First, we developed an overall model for the total sample in which the motivational variables were modelled as mediators between previous and subsequent mathematics achievement. Since models with latent motivational variables were computationally too complex, we worked with the manifest subscale scores of the

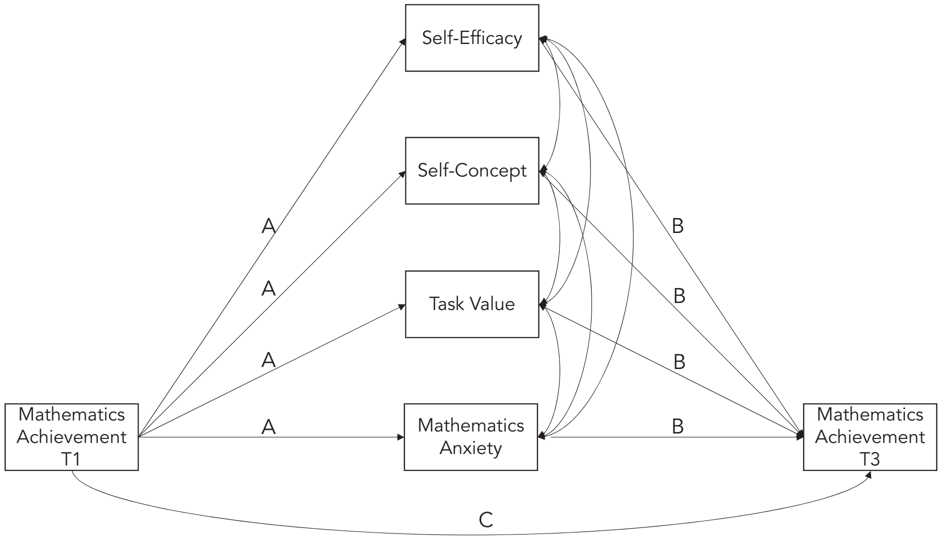


Figure 5.1 Conceptual model for Model 1 (basic mediation model).

motivation questionnaire. Figure 5.1 represents the basic mediation model which was tested in Model 1. Path *c* represents the stability effect from previous to subsequent mathematics achievement. The *a*-paths represent the effects from previous achievement on motivation. The *b*-paths represent the effects from motivation on subsequent achievement. The motivational variables mediate between previous and subsequent achievement if the indirect effect *ab* (the product of paths *a* and *b*) is significant. This model was further elaborated by adding working memory as a predictor of achievement by itself and in interaction with mathematics anxiety (to enhance readability, the exact sequence of models is described in the results section).

Second, the final overall model was estimated as a multiple-group model for three subsamples based on achievement at the T1 mathematics test. To create three approximately equally sized groups, students scoring more than half a standard deviation below the mean were assigned to the low-achieving subsample, students scoring more than half a standard deviation above the mean were assigned to the high-achieving subsample, and students scoring within half a standard deviation around the mean were assigned to the average-achieving subsample. To evaluate whether the relations between achievement and motivation were significantly different across the three achievement subsamples, Wald tests were performed. In order to limit the number of Wald tests (and diminish the probability of Type I errors), joint Wald tests ($df = 2$) which simultaneously compared all three achievement groups were used.

Following Hu and Bentler (1999), models with values above .95 for the comparative fit index (*CFI*) and Tucker-Lewis Index (*TLI*) and values below .06 and .08 for the root mean squared error of approximation (*RMSEA*) and the standardised root mean square residual (*SRMR*), respectively, were judged to have a good fit. In all models, predictor variables were grand-mean centered to facilitate interpretation of the results. Full information maximum likelihood estimation with robust standard errors was used to handle missing data and to correct for nonnormality. To correct for the nesting of students within classes, the *type=complex* option of M*plus* (which provides cluster-robust standard errors without building a full multilevel model) was used. To calculate indirect effects for the final models, the standardised parameter estimates of the *a* and *b* paths obtained in M*plus* were entered in the Rmediation package (Tofighi & MacKinnon, 2011) which provides 95% confidence intervals for the indirect effects based on the distribution of the product. The fully standardised mediated effect was used as an effect size measure for the indirect effect because of its satisfactory statistical properties and because it was of interest to evaluate the change in standard deviations of the outcome with a one standard deviation increase in the independent variable (Miočević, O'Rourke, MacKinnon, & Brown, 2017).

5.3 Results

Descriptive statistics are displayed in Table 5.1. The mean scores on the mathematics test indicate that the total sample and the average-achieving subsample scored close to the national average, whereas the subsamples of low-achieving and high-achieving students scored about one standard deviation below and above the mean, respectively. For all motivational variables, the low-achieving subsample scored lower and the high-achieving subsample scored higher compared to the average-achieving subsample. The distributions of the motivational variables were somewhat skewed, with relatively many high scores for self-efficacy, self-concept, and task value and relatively many low scores for mathematics anxiety (but the use of the maximum likelihood robust estimator corrected for this in the subsequent analyses).

Zero-order correlations between the variables are displayed in Table 5.2. All correlations were significant and in the expected direction. Self-efficacy and self-concept were strongly correlated (r = .80). Moreover, a combined perceived competence subscale representing the average of all self-efficacy and self-concept items had a very high internal consistency (α = .92). This raised the question whether these two constructs were empirically sufficiently distinct to be modelled as separate variables. In the subsequent analyses, we therefore compared a model in which the two variables were modelled separately to a model in which the variables were combined and proceeded with the best-fitting model.

Table 5.1 Descriptive statistics

	Total sample[a]	Low-achieving sub-sample[b]	Average-achieving sub-sample[c]	High-achieving sub-sample[d]			
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	Skewness (*SE*)	Kurtosis (*SE*)	% missing
Maths T1	0.11 (1.10)	-1.19 (0.62)	0.00 (0.28)	1.24 (0.67)	0.04 (0.04)	0.76 (0.08)	10.6
Maths T3	0.07 (1.13)	-1.03 (0.86)	0.02 (0.66)	0.99 (0.83)	-0.02 (0.04)	0.98 (0.08)	6.3
Self-efficacy	3.09 (0.55)	2.81 (0.56)	3.04 (0.50)	3.38 (0.44)	-0.37 (0.04)	0.07 (0.08)	4.6
Self-concept	3.05 (0.74)	2.56 (0.74)	3.01 (0.65)	3.50 (0.50)	-0.63 (0.04)	-0.24 (0.08)	4.6
Task value	3.00 (0.72)	2.82 (0.74)	2.99 (0.70)	3.17 (0.67)	-0.38 (0.04)	-0.72 (0.08)	4.6
Maths anxiety	1.67 (0.67)	1.95 (0.76)	1.67 (0.62)	1.42 (0.51)	1.07 (0.04)	0.65 (0.08)	4.6
Working memory	0.00 (0.16)	-0.06 (0.16)	0.01 (0.15)	0.05 (0.14)	-1.06 (0.04)	1.58 (0.08)	10.9

[a] $N = 4306$ [b] $N = 1064$ [c] $N = 1431$ [d] $N = 1358$. The achievement subsamples do not add up to the total sample size since students with missing data at the T1 mathematics test could not be assigned to an achievement subsample.

Table 5.2 Zero-order correlations

	Maths T1	Maths T3	Self-efficacy	Self-concept	Task value	Maths anxiety
Maths T1						
Maths T3	.80*					
Self-efficacy	.42*	.41*				
Self-concept	.52*	.51*	.80*			
Task value	.19*	.21*	.45*	.47*		
Maths anxiety	-.33*	-.32*	-.56*	-.60*	-.26*	
Working memory	.29*	.32*	.11*	.15*	.09*	-.08*

* $p < .01$

Model 1 tested the basic mediation model represented in Figure 5.1. Model 2 was identical to Model 1, except that self-efficacy and self-concept were combined into one variable labelled perceived competence. An overview of all models and their respective fit is provided in Table 5.3. Both Model 1 and Model 2 were saturated, but Model 2 yielded smaller values on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), indicating that it had a better fit than Model 1. Notably, in Model 1, self-concept significantly predicted subsequent math achievement (β = .123, *p* <.001), but self-efficacy did not (β = .001, *p* = .949), which can most probably be explained by multicollinearity of the two variables (see also Marsh et al., 2004). For the other parameters, Model 1 and Model 2 displayed similar results (detailed results for all models can be found in Appendix 5.1). Given the indications for a multicollinearity problem in Model 1 and the better fit of Model 2, we pursued the analyses with the combined perceived competence variable.

Table 5.3 Overview of models and model fit

Model	AIC	BIC	χ^2 (*df*), *p*	*RMSEA*	*CFI*	*TLI*	*SRMR*
1: as Figure 5.1	44720.765	44892.638	Saturated (*df* = 0)	-	-	-	-
2: self-efficacy and self-concept combined	40843.719	40971.032	Saturated (*df* = 0)	-	-	-	-
3: plus WM as predictor of maths T3	37129.589	37282.415	5.367 (3), .147	.014	.999	.997	.006
4: plus WM x maths anxiety interaction	30813.849	30998.514	30.23 (6), < .001	.031	.994	.983	.016
5: multiple-group as Model 3	27266.904	27717.380	13.834 (9), .128	.020	.998	.992	.012

WM = working memory

In Model 2, previous mathematics achievement had a strong positive effect on perceived competence (β = .506, *p* < .001) and a small to moderate positive effect on task value (β = .187, *p* < .001). Moreover, previous achievement had a moderate negative effect on mathematics anxiety (β = -.334, *p* < .001). Of the motivational variables, only perceived competence had a significant effect on subsequent achievement. This effect was positive as expected, but small in size (β = .115, *p* < .001). Previous achievement had a strong effect on subsequent achievement (β = .737, *p* < .001), indicating temporal stability.

In Model 3, working memory was added as a covariate. This model was no longer saturated and had a good fit. WM had a significant small positive effect on subsequent achievement (β = .092, *p* < .001). The other parameters in the model remained similar after controlling for WM.

In Model 4, an interaction term was added to test whether WM interacted with mathematics anxiety in predicting future achievement. The chi-square test of model fit was significant but this test is known to be oversensitive with large sample sizes, whereas model fit was good judging by the *RMSEA*, *CFI*, *TLI* and *SRMR*. However, the WM x mathematics anxiety interaction was not significant ($\beta = .005$, $p = .676$), indicating that the effect of mathematics anxiety did not vary at different levels of WM. Thus, mathematics anxiety had no significant effect on subsequent achievement, regardless of the WM level of the student. Therefore, Model 3 (excluding the interaction) was the final model for the full sample and its results are depicted in Figure 5.2. In Model 3, the estimated fully standardised indirect effect from previous achievement on subsequent achievement through perceived competence was 0.060 (*SE* = 0.009; p <.001; 95% *CI* = [0.043, 0.077]). This indicates that an increase of one standard deviation in previous achievement yields an increase of 0.060 standard deviation on subsequent achievement on the mathematics test through changes in perceived competence, which can be interpreted as a small indirect effect.

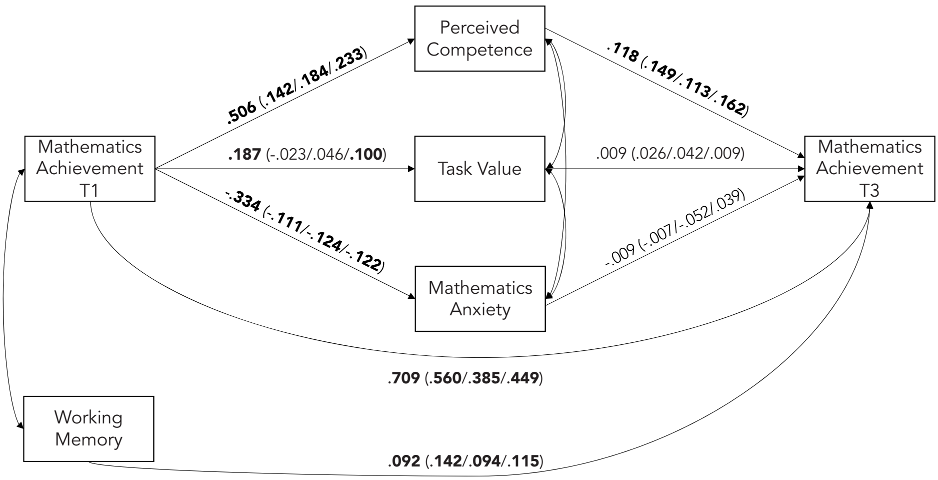


Figure 5.2 Final model results for the full sample (Model 3) and, in parentheses, for the achievement subsamples (low-achieving / average-achieving / high-achieving). Estimates printed in bold are significant (p < .01). Standard errors and estimated correlations were omitted from the figure but these can be found in Appendix 5.1.

In Model 5, the final overall model (i.e., analogous to Model 3) was estimated as a multiple-group model for the subsamples of low-achieving, average-achieving, and high-achieving students. This model had a good fit. Model results are depicted in Figure 5.2. For all achievement groups, previous achievement had a significant positive effect on perceived competence. However, the magnitude of the effect was significantly different across achievement groups (W [df = 2] = 18.13, p < .001), with relatively stronger effects for high-achieving students. The effect of previous achievement on task value also differed significantly across achievement groups (W [df = 2] = 7.05, p = .030) and was in fact only significant for high-achieving students. The Wald test also indicated that the effect of previous achievement on mathematics anxiety differed significantly across achievement groups (W [df = 2] = 8.08, p = .018). However, based on the parameter estimates the effect size seemed to be rather similar across achievement groups. This might be explained by the fact that the Wald test is not only based on effect size but also on the standard errors of the parameters and their covariance. Similar to the overall model, perceived competence was the only motivational variable with significant effects on subsequent achievement in all achievement groups. Although the parameter estimates indicated a potential difference between achievement groups with stronger effects for high-achieving students and weaker effects for average-achieving students, this difference did not reach significance (W [df = 2] = 4.74, p = .093). Given the large sample size, it seems safe to conclude that the effect from perceived competence on subsequent achievement did not differ substantially across achievement groups. The effects of task value on subsequent achievement was nonsignificant and similar across achievement groups (W [df = 2] = 0.49, p = .782). The effects of mathematics anxiety were not significant in any of the achievement groups and although the direction of effects differed between achievement groups, the Wald test indicated that between-group differences were not significant (W [df = 2] = 4.33, p = .109 for mathematics anxiety)). The estimated fully standardised indirect effects from previous achievement on subsequent achievement through perceived competence were 0.021 (SE = 0.009; p = .003; 95% CI = [0.007, 0.040]) for the low-achieving subsample, 0.021 (SE = 0.018; p = .005; 95% CI = [0.007, 0.037]) for the average-achieving subsample and 0.038 (SE = 0.011; p <.001; 95% CI = [0.021, 0.056]) for the high-achieving subsample.

5.4 Discussion

In this large-scale study, we examined the relations between several core aspects of motivation for mathematics – self-efficacy, self-concept, task value, and mathematics anxiety – and achievement on a standardised mathematics test. We investigated whether previous achievement predict that these motivational variables. Moreover, we examined how these motivational variables jointly predicted subsequent mathematics achievement and whether they mediated between previous and subsequent mathematics achievement. An innovative aspect of our study is that we also explored whether these relations were similar or different for students of diverse achievement levels within heterogeneous primary schools.

As expected, previous achievement had positive effects on perceived competence (self-efficacy and self-concept combined) and task value and negative effects on mathematics anxiety. This shows that, in line with former studies (e.g., Möller et al., 2014; Viljaranta et al., 2014), motivation is substantially related to previous achievement. This did not only hold for perceived competence – for which the effect may indicate a realistic self-perception based on previous achievement – but also for task value and mathematics anxiety. Thus, students with low previous performance were not only likely to feel less competent in mathematics, they generally also valued it less and experienced more mathematics anxiety. In fact, all motivational variables were moderately to strongly interrelated: Self-efficacy, self-concept and task value were positively related to each other and negatively related to mathematics anxiety. Indeed, the correlation between self-efficacy and self-concept was so strong that they had to be combined into a single perceived competence variable to circumvent multicollinearity problems. This contrasts with findings by Parker et al. (2014) in which self-efficacy and self-concept explained unique variance in achievement. However, this difference might be explained by the measures used, since Parker et al. used highly task-specific measures for self-efficacy (and not for self-concept), which were not feasible in the current study due to the broad age range (and according developmental level) of the participants.

Regarding the effects from the motivational variables on achievement, only perceived competence significantly predicted subsequent achievement. This corresponds with previous studies in which self-efficacy or self-concept had the largest predictive value (Bong et al., 2012; Kriegbaum et al., 2015; Spinath et al., 2006). Perceived competence partially mediated between previous and subsequent achievement, with a small but significant indirect effect. This is in line with previous findings in which self-efficacy mediated between previous and subsequent achievement for typically-achieving students (Jungert et al., 2014).

Task value and mathematics anxiety did not predict subsequent achievement, which might be explained by the inclusion of previous achievement as well as multiple motivational variables in the model. The zero-order correlations (reported in Table 5.2) showed that both task value (r = .21) and mathematics anxiety (r = -.31) were significantly related to subsequent achievement. However, these relations were no longer significant in the full model in which the effects of motivation on achievement were controlled for previous achievement and modelled besides the effects of other motivational variables. For task value, this lack of effects was in line with previous studies in which the effects of task value were controlled for previous achievement or modelled besides other motivational variables (Bouffard, Marcoux, Vezeau, & Bordeleau, 2003; Garon-Carrier et al., 2016; Spinath et al., 2006; although Kriegbaum et al. (2015) found a significant but very small effect in secondary school). This does not necessarily mean that task value has no effect on subsequent achievement when modelled as a single predictor, since the substantial correlation between task value and perceived competence may cause a reciprocal suppression effect (Plante, De la Sablonnière, Aronson, & Théorêt, 2013). As explained by Plante et al. (2013), it is possible that the variance which task value shares with perceived competence has a positive effect on subsequent achievement, while the unique variance of task value which does not overlap with perceived competence has no or even a negative effect on achievement. For example, high task value might be stressful when it is not coupled with high perceived competence: students with this combination might feel the need to achieve well in mathematics, but might not feel able to do so. The findings of our study and previous studies (Bouffard et al., 2003; Garon-Carrier et al., 2016; Plante et al., 2013; Spinath et al., 2006) suggest that valuing a subject is not, by itself, enough to enhance achievement. We speculate that, while task value may trigger students to engage in an activity, it needs to be coupled with perceived competence (which may arise over time if the student experiences success during the activity) to potentially enhance achievement.

Regarding mathematics anxiety, the results were not in line with our hypothesis that mathematics anxiety would predict subsequent achievement (perhaps only for students with high WM capacity). However, this hypothesis was based on previous studies that typically did not control for previous achievement and did not include other motivational variables in the model. When perceived competence and mathematics anxiety are simultaneously modelled as predictors of subsequent achievement, a similar reciprocal suppression effect might come into play. The unique variance of mathematics anxiety which does not overlap with perceived competence might be interpreted as a student's general inclination towards anxiety, regardless of the student's perceived competence in mathematics. Perhaps, this anxiety is not very harmful as long as it is not coupled with low perceived competence, since students with high perceived competence might have little reason to be anxious specifically for mathematics. Recent research suggests that mathematics anxiety is an effect of low perceived competence (Van der Beek et al., 2017). Another possible explanation for the lack of effects of mathematics anxiety on subsequent achievement in the current study is that we used a longitudinal design with control for previous achievement. In contrast, most previous studies about mathematics anxiety used cross-sectional designs in which mathematics anxiety and achievement were measured at the same timepoint. Perhaps, for anxiety, concurrent effects could be more relevant than longitudinal effects, since anxiety is supposed to interfere directly with performance during testing. Previous longitudinal studies (Krinzinger et al., 2009; Vukovic et al., 2013) also found no general effect of mathematics anxiety on subsequent achievement after controlling for previous achievement. Vukovic et al. (2013) did find a longitudinal effect, but only on application problems and only for students with high WM capacity. In our study, this interaction with WM was not replicated (despite the use of application problems in the mathematics achievement test). More research is necessary to clarify in which situations mathematics anxiety interacts with WM.

Regarding the moderation of the investigated effects by achievement level, we found that only the effects of previous achievement on motivation (and not of motivation on subsequent achievement) differed between achievement groups. For the interpretation of these effects, it should be kept in mind that the multiple-group analysis evaluated these effects *within* achievement groups. Thus, the effects from achievement on motivation were now based on achievement relative to the homogeneous subsample (low-achieving, average-achieving or high-achieving) rather than relative to the total heterogeneous sample. Therefore, the variation in initial achievement level was smaller and this probably explains why the effects within the subsamples were generally smaller than the overall effects. Nevertheless, the effect of previous achievement on perceived competence was significant and positive for students of all achievement levels. The strength of this effect significantly differed between achievement groups, with a larger effect for high-achieving students. Perhaps, high-achieving students are better able to construct realistic perceptions of their own abilities based on previous achievement. An alternative possibility is that an evaluation of one's own achievement relative to students of a similar achievement level is more relevant or salient for high-achieving students than for low-achieving students (e.g., it could be that low-achieving students primarily compare their own achievement to the achievement of the whole class rather than to other low-achieving students, whereas high-achieving students might compare themselves more often to other high-achieving students). Our study does not inform about these processes, but these are interesting hypotheses to explore in future research.

For task value, the effect of previous achievement was again strongest for high-achieving students. In fact, these effects were not significant in the subsamples of low-achieving and average-achieving students, indicating that variations in achievement level within the groups of low-achieving and average-achieving students did not influence the degree to which they valued mathematics. This seems to contrast with the findings of Jõgi et al. (2015) in which the effects of previous achievement on task value were stronger for low-ability students. However, these results cannot be directly compared since the study by Jõgi et al. assigned students to subsamples based on general cognitive ability (rather than mathematics achievement) and did not distinguish between average-ability and high-ability students. In the current study, high-achieving students seemed to be more sensitive to previous success (or failure): besides the effect on perceived competence, previous achievement relative to other high-achieving students also had an effect on students' enthusiasm for mathematics. Future research could investigate whether this difference can be replicated with other achievement measures (e.g., grades given by the teacher) and explore possible mechanisms behind this difference. For mathematics anxiety, the effects of previous achievement did not show meaningful differences between achievement groups, with a small but significant effect on mathematics anxiety in each group. Thus, even within the relatively homogeneous subsamples of low-achieving, average-achieving, and high-achieving students, individual differences in mathematics achievement were related to mathematics anxiety. This might indicate that students compare themselves not only to the whole class but also to subgroups of students with a similar achievement level. Even for high-achieving students, for example, relatively low previous achievement compared to other high-achieving students (which would typically still be high achievement compared to the class average) might provoke mathematics anxiety.

The effects in the opposite direction – i.e., from motivation on subsequent achievement – were similar across achievement groups. Mathematics anxiety and task value did not have a significant effect on subsequent achievement in any of the achievement groups. Thus, it is not the case that these aspects of motivation are only related to subsequent achievement in one particular achievement group (which might have been obscured in the analysis of the total sample). For all achievement groups, perceived competence was the only motivational variable with a significant effect on subsequent achievement. Thus, also within the relatively homogeneous achievement groups, students with higher perceived competence subsequently achieved more highly after controlling for individual differences in previous achievement, while this was not the case for mathematics anxiety and task value. These findings mirror the findings in the total sample, and the same potential explanations for the lack of effects of task value and mathematics anxiety may apply (control for previous achievement and the inclusion of multiple motivational variables with shared variance). The effect of perceived competence was small and similar across achievement groups. This finding contrasts with one study by Jungert et al. (2014), in which this effect was significant only for typically-achieving students and not for low-achieving students. However, the relatively small sample size of that study may have limited the power to detect effects.

5.4.1 Limitations, implications and directions for future research

The results of the current study should be considered in light of the following limitations. First, the timeframe of the study was limited to three measurement occasions spread over half a year. The processes within the timeframe of our study are likely to be influenced as well by educational experiences before the beginning of the study. Especially in the higher grades, previous achievement is confounded with previous educational quality and motivation for mathematics. A related limitation is that we only measured one construct at each time point (T1: achievement – T2: motivation – T3: achievement). Thus, the effects of previous achievement on motivation were not controlled for previous motivation. Future studies could follow students from the very beginning of (preparatory) mathematics education onwards and use cross-lagged designs to track the development of students' motivation and achievement over a longer period of time. This would enable stronger causal inferences. On the other hand, our design was very suitable to test mediation due to the temporal ordering of the measures. A second limitation is that we did not investigate *how* achievement and motivation might influence each other. Future research could spend more attention on these processes, for example using behavioural data on potentially relevant variables such as time on task. However, this may require more intensive research methods – e.g., experimental manipulation or classroom observations – which are more suitable for smaller samples. In contrast, the large scale of the current study is one of its strengths, because this enhances the generalisability of the findings. A third limitation is that mathematics achievement was the only outcome variable. While the use of a standardised, nationally administered achievement test is a strength of this study, future research could also consider other outcomes. For example, it is conceivable that task value has positive effects on other outcomes (e.g., choosing mathematics as a subject in secondary school) which do not directly translate into achievement on a standardised test.

Despite these limitations, the results of our study have the following implications. Motivation is substantially related to previous achievement. This holds for all aspects of motivation that were investigated in the current study. After controlling for previous achievement, the relations between motivation and subsequent achievement were less pronounced. Mathematics anxiety and task value were not significantly related to subsequent achievement, neither in the total sample nor in the subsamples. Perceived competence did have a small but significant effect on subsequent achievement after controlling for previous achievement and working memory, both in the total sample and within each achievement group. Thus, the relative importance of diverse aspects of motivation seems to be similar across achievement groups. These results provide no indications that specific aspects of motivation should be differentially fostered depending on students' achievement level. Rather, the positive effect of perceived competence across achievement groups implies that fostering perceived competence might be a promising way to enhance achievement for students of all achievement levels. However, this conclusion should be drawn with care since our results also indicate that actual achievement is an important source of perceived competence. Thus, the effects seem to be reciprocal. It does not seem to make sense to foster unrealistically high perceived competence, but it might be helpful to create situations in which students can experience mastery. This might enhance students' perceived competence and, in turn, adaptive learning behaviours and subsequent achievement. Interestingly, in a study in which the success rate of solving mathematics problems was experimentally manipulated (Jansen et al., 2013), students who experienced more success did not show higher perceived competence (contrary to expectations) but did attempt to solve more problems and increased more in performance compared to students who experienced less success. One opportunity for experiencing success might be to give specific mastery-oriented feedback when a student performs a task successfully ("You really know how to solve this kind of sums now, don't you?"). Another opportunity might be to adapt the challenge level of mathematics tasks to students' current level of understanding to ensure that the tasks are challenging but realistic for all students. One currently popular way to enable all students to work in their zone of proximal development (in which success is just within reach) is differentiated instruction, an educational approach in which goals, instruction and practice are adapted to students' educational needs based on current achievement level (Prast et al., 2015; Roy, Guay, & Valois, 2013). In this way, especially low-achieving students might experience less failure and more success, which might have positive effects on perceived competence. On the other hand, low-achieving students might start to feel less confident about their own competence when they are aware that they receive additional instruction or work on easier tasks than their peers. When differentiated instruction is used, it seems to be important that teachers are aware of such social comparison processes and attempt to reduce their potential negative effects – for example by promoting a positive attitude towards diversity and a dynamic rather than fixed view of students' abilities (Dweck, 2000; Tomlinson et al., 2003). An alternative possibility suggested by Roy, Guay, and Valois (2015) is that differentiated instruction reduces social comparison effects (rather than amplifying them) because the focus on the individual learning process may promote an internal rather than external frame of reference. Indeed, these authors found that the big-fish-little-pond effect (a negative effect of high class-average achievement on individual self-concept) was less pronounced for low-achieving students if teachers frequently used differentiated instruction (Roy et al., 2015). Clearly, the effects of differentiated instruction on perceived competence are an interesting topic for future research. Specifically, future research could investigate whether potential negative effects for low-achieving students can be overcome while preserving potential positive effects of differentiated instruction on perceived and actual mathematical competence.

A general conclusion from the current study is that motivation and achievement seem to be intertwined. The various aspects of motivation are also related to each other. Our results imply that a powerful way to increase motivation may be to help students to master their mathematics tasks – for example by providing excellent instruction and ample opportunities for practice, adapted to students' current level of understanding where necessary. This may enhance both achievement and perceived competence, thus starting a positive cycle of reciprocal effects.

References

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*, 20-29.

Alloway, T., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology*, *28*, 725-734. doi:10.1080/01443410802243828

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, *11*, 181-185. doi:10.1111/1467-8721.00196

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191-215. doi:10.1037/0033-295X.84.2.191

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Birgin, O., Baloğlu, M., Çatlıoğlu, H., & Gürbüz, R. (2010). An investigation of mathematics anxiety among sixth through eighth grade students in Turkey. *Learning and Individual Differences*, *20*, 654-658. doi:10.1016/j.lindif.2010.04.006

Bong, M., Cho, C., Ahn, H. S., & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *Journal of Educational Research*, *105*, 336-352. doi:10.1080/00220671.2011.627401

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*, 1-40. doi:1040-726X/03/0300-0001/0

Bouffard, T., Marcoux, M. F., Vezeau, C., & Bordeleau, L. (2003). Changes in self-perceptions of competence and intrinsic motivation among elementary school children. *British Journal of Educational Psychology*, *73*, 171-186. doi:10.1348/00070990360626921

Chen, S. K., Yeh, Y. C., Hwang, F. M., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort-multioccasion study. *Learning and Individual Differences*, *23*, 172-178. doi:10.1016/j.lindif.2012.07.021

Corpus, J. H., McClintic-Gilbert, M. S., & Hayenga, A. O. (2009). Within-year changes in children's intrinsic and extrinsic motivational orientations: Contextual predictors and academic outcomes. *Contemporary Educational Psychology*, *34*, 154-166. doi:10.1016/j.cedpsych.2009.01.001

Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I'm able, and I know I am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development*, *78*, 430-447. doi:10.1111/j.1467-8624.2007.01007.x

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Fast, L. A., Lewis, J. L., Bryant, M. J., Bocian, K. A., Cardullo, R. A., Rettig, M., & Hammond, K. A. (2010). Does math self-efficacy mediate the effect of the perceived classroom environment on standardized math test performance? *Journal of Educational Psychology*, *102*, 729-740. doi:10.1037/a0018863

Friso-Van den Bos, I., Van der Ven, S. H. G., Kroesbergen, E. H., & Van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, *10*, 29-44. doi:10.1016/j.edurev.2013.05.003

Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J. -., ... Tremblay, R. E. (2016). Intrinsic motivation and achievement in mathematics in elementary school: A longitudinal investigation of their association. *Child Development*, *87*, 165-175. doi:10.1111/cdev.12458

Gniewosz, B., Eccles, J. S., & Noack, P. (2015). Early adolescents' development of academic self-concept and intrinsic task value: The role of contextual feedback. *Journal of Research on Adolescence*, *25*, 459-473. doi:10.1111/jora.12140

Hampton, N. Z., & Mason, E. (2003). Learning disabilities, gender, sources of efficacy, self-efficacy beliefs, and academic achievement in high school students. *Journal of School Psychology*, *41*, 101-112. doi:10.1016/S0022-4405(03)00028-1

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, *21*, 33-46. doi:10.2307/749455

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55. doi:10.1080/10705519909540118

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*, 505-528. doi:10.1016/j.jsp.2011.07.001

Huang, C. (2012). Discriminant and incremental validity of self-concept and academic self-efficacy: A meta-analysis. *Educational Psychology*, *32*, 777-805. doi:10.1080/01443410.2012.732386

Hughes, A., Galbraith, D., & White, D. (2011). Perceived competence: A common core for self-efficacy and self-concept? *Journal of Personality Assessment*, *93*, 278-289. doi:10.1080/00223891.2011.559390

Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H. G., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, *24*, 190-197. doi:10.1016/j.lindif.2012.12.014

Janssen, J., Scheltens, F., & Kraemer, J. M. (2005). *Leerling- en onderwijsvolgsysteem rekenen-wiskunde* [Student progress monitoring system mathematics]. Arnhem, The Netherlands: Cito.

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS rekenenwiskunde voor groep 3 tot en met 8* [Scientific justification of the mathematics test for grade 1 through grade 6]. Arnhem, The Netherlands: Cito.

Jõgi, A., Kikas, E., Lerkkanen, M., & Mägi, K. (2015). Cross-lagged relations between math-related interest, performance goals and skills in groups of children with different general abilities. *Learning and Individual Differences*, *39*, 105-113. doi:10.1016/j.lindif.2015.03.018

Jungert, T., Hesser, H., & Träff, U. (2014). Contrasting two models of academic self-efficacy – domain-specific versus cross-domain – in children receiving and not receiving special instruction in mathematics. *Scandinavian Journal of Psychology*, *55*, 440-447. doi:10.1111/sjop.12139

Keuning, J., Van Boxtel, H., Lansink, N., Visser, J., Weekers, A., & Engelen, R. (2015). *Actualiteit en kwaliteit van normen: Een werkwijze voor het normeren van een leerlingvolgsysteem* [Up-to-dateness and quality of norms: A method to develop norms for a student progress monitoring system]. Arnhem, The Netherlands: Cito.

Kriegbaum, K., Jansen, M., & Spinath, B. (2015). Motivation: A predictor of PISA's mathematical competence beyond intelligence and prior test achievement. *Learning and Individual Differences*, *43*, 140-148. doi:10.1016/j.lindif.2015.08.026

Krinzinger, H., Kaufmann, L., & Willmes, K. (2009). Math anxiety and math ability in early primary school years. *Journal of Psychoeducational Assessment*, *27*, 206-225. doi:10.1177/0734282908330583

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, *30*, 520-540. doi:10.2307/749772

Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: A reexamination of relations between self-efficacy, self-concept, and achievement. *Journal of Educational Psychology*, 96, 518-522. doi:10.1037/0022-0663.96.3.518

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. British Journal of Educational Psychology, 81, 59-77. doi:10.1348/000709910X503501

Miočević, M., O'Rourke, H. P., MacKinnon, D. P., & Brown, H. C. (2017). Statistical properties of four effect-size measures for mediation models. *Behavior Research Methods*, Epub ahead of print doi:10.3758/s13428-017-0870-1.

Möller, J., Zimmermann, F., & Köller, O. (2014). The reciprocal internal/external frame of reference model using grades and test scores. *British Journal of Educational Psychology*, *84*, 591-611. doi:10.1111/bjep.12047

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology*, *106*, 1170-1191. doi:10.1037/a0036307

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, *34*, 29-48. doi:10.1080/01443410.2013.797339

Plante, I., De la Sablonnière, R., Aronson, J. M., & Théorêt, M. (2013). Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology*, *38*, 225-235. doi:10.1016/j.cedpsych.2013.03.004

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2012). *Handleiding voor de Rekenmotivatievragenlijst voor Kinderen* [Manual for the Mathematics Motivation Questionnaire for Children]. Utrecht, The Netherlands: Utrecht University.

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2015). Differentiation in primary school mathematics: Expert recommendations and teacher self-assessment. *Frontline Learning Research*, *3*, 90-116. doi:10.14786/flr.v3i2.163

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2018). Differentiated instruction in primary mathematics: Effects of teacher professional development on student achievement. *Learning and Instruction*, *54*, 22-34, doi: 10.1016/j.learninstruc.2018.01.009

Preckel, F., Niepel, C., Schneider, M., & Brunner, M. (2013). Self-concept in adolescence: A longitudinal study on reciprocal effects of self-perceptions in academic and social domains. *Journal of Adolescence*, *36*, 1165-1175. doi:10.1016/j.adolescence.2013.09.001

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology*, *141*, 83-100. doi:10.1016/j.jecp.2015.07.014

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, *14*, 187-202. doi:10.1080/15248372.2012.664593

Roy, A., Guay, F., & Valois, P. (2013). Teaching to address diverse learning needs: Development and validation of a differentiated instruction scale. *International Journal of Inclusive Education*, *17*, 1186-1204. doi:10.1080/13603116.2012.743604

Roy, A., Guay, F., & Valois, P. (2015). The big-fish–little-pond effect on academic self-concept: The moderating role of differentiated instruction and individual achievement. *Learning and Individual Differences*, *42*, 110-116. doi:10.1016/j.lindif.2015.07.009

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*, 407-441. doi:10.3102/00346543046003407

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, *34*, 363-374. doi:10.1016/j.intell.2005.11.004

Suinn, R. M., & Winston, E. H. (2003). The mathematics anxiety rating scale, a brief version: Psychometric data. *Psychological Reports*, *92*, 167-173. doi:10.2466/pr0.2003.92.1.167

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, *43*, 692-700. doi:10.3758/s13428-011-0076-x

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., ... Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, *27*, 119-145. doi:10.1177/016235320302700203

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*, 111-133. doi:10.1207/s15326985ep3902_3

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Prast, E. J., & Van Luit, J. E. H. (2015). Validity and reliability of an online working memory task for self-reliant administration in school-aged children. *Behavior Research Methods*, *47*, 708-719. doi:10.3758/s13428-014-0469-8

Van der Beek, J. P. J., Van der Ven, S. H. G., Kroesbergen, E. H., & Leseman, P. P. M. (2017). Self-concept mediates the relation between achievement and emotions in mathematics. *British Journal of Educational Psychology*, *87*, 478-495. doi:10.1111/bjep.12160

Viljaranta, J., Tolvanen, A., Aunola, K., & Nurmi, J. E. (2014). The developmental dynamics between interest, self-concept of ability, and academic performance. *Scandinavian Journal of Educational Research*, *58*, 734-756. doi:10.1080/00313831.2014.904419

Von Maurice, J., Dörfler, T., & Artelt, C. (2014). The relation between interests and grades: Path analyses in primary school age. *International Journal of Educational Research*, *64*, 1-11. doi:10.1016/j.ijer.2013.09.011

Vukovic, R. K., Kieffer, M. J., Bailey, S. P., & Harari, R. R. (2013). Mathematics anxiety in young children: Concurrent and longitudinal associations with mathematical performance. *Contemporary Educational Psychology*, *38*, 1-10. doi:10.1016/j.cedpsych.2012.09.001

Wang, Z., Lukowski, S. L., Hart, S. A., Lyons, I. M., Thompson, L. A., Kovas, Y., ... Petrill, S. A. (2015). Is math anxiety always bad for math learning? The role of math motivation. *Psychological Science*, *26*, 1863-1876. doi:10.1177/0956797615602471

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, *30*, 1-35. doi:10.1016/j.dr.2009.12.001

Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68-81. doi:10.1006/ceps.1999.1015

Wu, S. S., Barth, M., Amin, H., Malcarne, V., & Menon, V. (2012). Math anxiety in second and third graders and its relation to mathematics achievement. *Frontiers in Psychology*, *3*, 1-11. doi: 10.3389/fpsyg.2012.00162

Appendix 5.1
Detailed results for all models

Table A5.1 Standardised results of Model 1 with self-efficacy and self-concept as separate variables (*N* = 4297)

Parameter	Estimate (SE)
Predictors	
T1 Achievement → Self-efficacy	.425** (.019)
T1 Achievement → Self-concept	.524** (.017)
T1 Achievement → Task value	.186** (.019)
T1 Achievement → Mathematics anxiety	-.334** (.014)
T1 Achievement → T3 Achievement	.732** (.013)
Self-efficacy → T3 Achievement	.001 (.019)
Self-concept → T3 Achievement	.123** (.020)
Task value → T3 Achievement	.011 (.013)
Mathematics anxiety → T3 Achievement	-.007 (.013)
Correlations	
Self-efficacy with Self-concept	.755** (.008)
Self-efficacy with Task value	.413** (.017)
Self-efficacy with Mathematics anxiety	-.493** (.014)
Self-concept with Task value	.449** (.018)
Self-concept with Mathematics anxiety	-.529** (.014)
Task value with Mathematics anxiety	-.212** (.018)

** $p < .001$.
For parsimony, means (close to 0 due to centering) and variances are omitted from the table.

Table A5.2 Standardised results of Models 2 through 5

	Model 2 N = 4297 Estimate (SE)	Model 3 N = 4306 Estimate (SE)	Model 4 N = 4306 Estimate (SE)	Model 5 Low-achieving n = 1064 Estimate (SE)	Model 5 Average-achieving n = 1431 Estimate (SE)	Model 5 High-achieving n = 1358 Estimate (SE)
Predictors						
T1 Ach. → PC	.506** (.018)	.506** (.018)	.505** (.018)	.142* (.043)	.184** (.025)	.233** (.026)
T1 Ach. → TV	.187** (.019)	.187** (.019)	.187** (.019)	-.023 (.036)	.046 (.026)	.100** (.028)
T1 Ach. → MA	-.334** (.014)	-.334** (.014)	-.333** (.014)	-.111* (.032)	-.124** (.027)	-.122** (.027)
T1 Ach. → T3 Ach.	.737** (.013)	.709** (.014)	.709** (.014)	.560** (.034)	.385** (.026)	.449** (.026)
WM → T3 Ach.	NA	.092** (.013)	.092** (.013)	.142** (.032)	.094* (.029)	.115** (.024)
PC → T3 Ach.	.115** (.017)	.118** (.017)	.118** (.017)	.149** (.038)	.113* (.038)	.162** (.034)
TV → T3 Ach.	.013 (.013)	.009 (.013)	.009 (.013)	.026 (.035)	.042 (.028)	.009 (.028)
MA → T3 Ach.	-.009 (.013)	-.009 (.013)	-.009 (.013)	-.007 (.035)	-.052 (.032)	.039 (.029)
WM x MA → T3 Ach.	n/a	n/a	.005 (.011)	n/a	n/a	n/a
Correlations						
PC with TV	.462** (.017)	.462** (.017)	.462** (.017)	.450** (.032)	.505** (.024)	.402** (.029)
PC with MA	-.547** (.013)	-.548** (.013)	-.548** (.013)	-.515** (.021)	-.564** (.025)	-.524** (.024)
TV with MA	-.212** (.018)	-.212** (.018)	-.212** (.018)	-.131** (.033)	-.258** (.029)	-.201** (.030)
T1 Ach. with WM	NA	.300** (.023)	.300** (.023)	.184** (.043)	.088* (.026)	.120** (.032)

* $p < .01$ ** $p < .001$
T1 Ach. = mathematics achievement at T1, T3 Ach. = mathematics achievement at T3, PC = perceived competence, TV = task value, MA = mathematics anxiety, WM = working memory. For parsimony, means (close to 0 due to centering) and variances are omitted from the table.

6

Summary and general discussion

Adapting education to the diverse needs of students with a broad range of academic ability and achievement levels is a challenge for teachers and for the field of teaching and teacher education in general. Differentiation for students of diverse achievement levels, also called readiness-based differentiation or cognitive differentiation, was the central theme of this dissertation. Since differentiation is grounded in a particular content domain (Vogt & Rogalla, 2009), this dissertation zoomed in on one content domain, namely mathematics education in primary school. The achievement of Dutch students in mathematics has been decreasing in international comparisons and a need for teacher professional development about differentiation in mathematics has been identified (KNAW, 2009; Meelissen & Punter, 2016).

The first goal of this dissertation was to specify what differentiation entails in the context of primary mathematics education. The second goal was to investigate the degree to which teachers implement the various recommended strategies for readiness-based differentiation in primary school mathematics education. The third goal was to develop and evaluate a teacher professional development (PD) programme about differentiation in primary school mathematics. The fourth goal was to investigate the reciprocal relations between achievement and several aspects of motivation for mathematics both in general and for subsamples of low-achieving, average-achieving, and high-achieving students.

The data on which this dissertation is based were collected in the context of project GROW (Gedifferentieerd RekenOnderWijs; a Dutch acronym for differentiated primary mathematics education). In the first phase of the project, an expert consensus procedure was used to specify what differentiated primary mathematics education entails (goal N° 1, Chapter 2). Based on this specification, a PD programme about differentiation in primary mathematics education was developed. In the second phase of the project, this PD programme was implemented and evaluated in a large-scale study involving 32 schools (see Figure 1.1 in Chapter 1 for the design of the study). Both teacher-level data and student-level data were collected over two years. These data were used to examine teachers' implementation of differentiation (goal N° 2, Chapter 2 and 3), to evaluate the effects of the PD programme on teachers' behaviour and student achievement (goal N° 3, Chapter 3 and 4), and to investigate the relations between motivation and achievement in mathematics for students of diverse achievement levels (goal N° 4, Chapter 5). An overview of the goals and chapters of this dissertation can be found in Figure 1.2 (Chapter 1). In this final chapter, the main findings are summarised and limitations and implications are discussed in relation to each goal. General reflections and directions for future research are subsequently provided, followed by the main conclusions and implications.

6.1 Goal N° 1: Specification of differentiation in mathematics

As has been argued in Chapter 2 of this dissertation, differentiation is an umbrella term which has been used to refer to a broad range of educational interventions. Therefore, an expert consensus procedure was used to specify what readiness-based differentiation entails in the context of primary mathematics education. The expert consensus procedure resulted in a five-step model for differentiation called the cycle of differentiation (see Figure 2.2 in Chapter 2). The first step in the cycle of differentiation is the identification of educational needs. Based on an analysis of students' current skill level, the teacher should divide the students over within-class achievement groups (typically low-achieving, average-achieving, and high-achieving). The grouping arrangement should be flexible, enabling students to switch between groups based on changes in their educational needs. In addition to formal measures such as achievement tests, informal measures such as the analysis of daily work and diagnostic conversations should be used on an ongoing basis to signal changes in educational needs and to determine educational needs in a more qualitative and refined way (e.g., if a student struggles with a particular type of sums, why is this the case and what does the student need to overcome this problem?). In the second step of the cycle, the teacher should set differentiated goals which should be challenging but realistic for the students in the different subgroups. In the third step, the teacher should differentiate instruction by providing broad whole-class instruction which engages students of diverse achievement levels, by providing subgroup instruction tailored to the needs of that subgroup, and by making adaptations for individual students. In the fourth step, the practice tasks should be differentiated both quantitatively and qualitatively for low-achieving and high-achieving students. In the fifth and final step, the teacher should evaluate both the learning progress (i.e., whether the students have met the goals) and the learning process (i.e., whether the applied adaptations of instruction and practice were effective). The evaluation phase informs the teacher about students' current achievement level and about instructional approaches that work for these students, completing the cycle and serving as new input for the identification of educational needs. Since attention for organisational aspects (e.g., planning and classroom management) is required in the implementation of each step of the cycle of differentiation, organisation is placed centrally in the cycle.

The steps in the cycle of differentiation are in line with the two core components of differentiated instruction identified by Roy, Guay and Valois (2013): progress monitoring (step 1 and 5) and instructional adaptations (step 2, 3 and 4). An innovative aspect of the study described in Chapter 2 is that the experts did not only achieve consensus on which steps should be included in the model for differentiation, but also agreed on *how* progress should be monitored and *how* goals, instruction and practice should be adapted to the educational needs of students with diverse achievement levels. For each step in the cycle of differentiation, a set of strategies for differentiation was recommended (see Chapter 2 for a detailed description of the cycle of differentiation and the recommended strategies for each step).

A limitation of this study is that consensus procedures are inherently restricted to the participating experts. The risk that other experts might have provided different input cannot be eliminated but was diminished in this study by including experts of several different institutions for both pre-service and in-service teacher training. In our experience, the consensus procedure – a combination of the Delphi method and focus group discussions – was an effective way to collect the experiential knowledge of the participating experts systematically. Moreover, these pre-service and in-service teacher educators served as a valuable bridge between the theory and practice of teaching. The results of the study advance the field by providing a prescriptive model for differentiation as well as practical strategies for each step.

The model resulting from this study also illustrates the complexity of implementing differentiation. Teachers should be able to diagnose students' educational needs, set differentiated goals, differentiate instruction and practice, evaluate students' learning progress and process, and they should have good organisational skills to implement and coordinate these aspects of differentiation. All steps of the cycle are interdependent: For example, appropriate differentiation of instruction relies on an accurate diagnosis of educational needs and appropriately matched learning goals. These results underline the importance of all four dimensions of adaptive teaching competency - subject matter knowledge, the ability to diagnose students' current understanding and achievement, the ability to use diverse teaching methods to meet diverse students' needs, and classroom management skills (Vogt & Rogalla, 2009) - and also raise the question to what extent teachers implement the recommended strategies for differentiation.

6.2 Goal N° 2: Implementation of differentiation

To investigate the degree to which teachers implemented the various recommended strategies for readiness-based differentiation in mathematics, two newly developed measures were used. A self-report questionnaire was administered in the total sample of teachers (reported in Chapter 2) and video observations were carried out in a subsample (reported in Chapter 3). Taken together, the results of the self-report questionnaire and the video observations provided the following insights into the implementation of differentiation by Dutch primary school teachers. Already at baseline (i.e., before the start of the PD programme in any of the cohorts), most teachers worked with homogeneous within-class ability groups. However, differentiation entails more than working with ability groups: It is especially relevant whether and how the ability groups are used to match instruction and practice to students' educational needs. Both the self-report and observational data indicated that different kinds of adaptations were made for low-achieving vs. high-achieving students. For low-achieving students, the focus was on differentiation of instruction, often by providing extended instruction in a subgroup. Differentiation of the practice tasks was also used occasionally, especially if the curriculum provided differentiated tasks for low-achieving students. For high-achieving students, the focus was on differentiation of the practice tasks, mostly using enrichment tasks provided by the curriculum. In contrast, subgroup instruction for high-achieving students was almost never observed. These findings place previously expressed concerns regarding a lack of differentiation for high-achieving students (Brighton, Moon, & Huang, 2015; Hertberg-Davis, 2009) in a different perspective. Many teachers do implement some differentiation for high-achieving students, indicating that they do perceive a need for differentiation for these students, but they focus on differentiation of the practice tasks rather than on differentiation of instruction. However, this practice should be critically evaluated. Assigning ready-made enrichment tasks provided by the curriculum without further consideration may be an easy way to keep high-achieving students busy, but is no guarantee for meeting the specific educational needs of these students. For example, it is not always clear whether these pre-differentiated tasks provided by the curriculum have an appropriate challenge level for the high-achieving students at hand (note that the appropriate challenge level may also differ substantially *between* high-achieving students within one classroom). Furthermore, high-achieving students also need at least some instruction or guidance when working on sufficiently challenging enrichment tasks (VanTassel-Baska & Stambaugh, 2005). The lack of specific instruction for high-achieving students may reflect the misconception that high-achieving students do not need instruction or guidance but may also result from low self-efficacy for providing subgroup instruction at a high challenge level in mathematics or from a (perceived) lack of time to attend to these students (Rubenstein, Gilson, Bruce-Davis, & Gubbins, 2015; Van de Weijer-Bergsma et al., 2016; VanTassel-Baska & Stambaugh, 2005). Future research could examine these potential barriers, as well as ways to overcome them, in more detail.

Another finding that emerged from both the self-report questionnaire and the observational data is that teachers were more likely to use general strategies for differentiation that could be implemented relatively easily for all students (e.g., using multiple modalities), whereas specialised strategies aimed at the specific needs of low-achieving or high-achieving students were used less frequently. Implementing specialised strategies – for example, the use of diagnostic conversations to determine individual educational needs – may not only require more time, but also more knowledge and skills (McLeskey & Waldron, 2011). While teachers frequently provided additional subgroup instruction to low-achieving students, the quality of this additional instruction could be improved. Teachers occasionally used recommended strategies for students who experience difficulties in understanding (certain aspects of) mathematics, such as teaching at a lower level of abstraction and allowing students to work with manipulatives. When such adaptations were observed, teachers often followed the suggestions for additional instruction as provided in the teacher manual of the curriculum, although other teachers also used this type of adaptations on their own initiative. However, systematic attempts to find out *why* a student could not solve a particular type of sums followed by an appropriate instructional adaptation were hardly ever observed. Note that this is a qualitative observation, since the observation instrument focused on instructional adaptations and not on the process of diagnosing students' educational needs.

A strong asset of this study was that the implementation of differentiation in mathematics for both low-achieving and high-achieving students could be investigated in more detail than in previous studies due to the use of two new measures developed specifically for this purpose. While the use of newly developed measures enables innovative research, a drawback is that the reliability and validity of new instruments are less established. In addition, both types of methodologies (self-report and observation) have their own strengths and limitations (Desimone, 2009). Self-report instruments may be biased by a tendency to give socially desirable answers. Observational instruments may be more objective in the sense that they are rated by external observers, but the standards for what was rated as low or high usage in the current observation instrument were not based on an established gold standard (in contrast to well-established criterion-referenced observation instruments for other aspects of teaching such as the Classroom Assessment Scoring System; Pianta, La Paro, & Hamre, 2008). Rather, they were newly developed based on the input from the experts in the consensus procedure as well as on what we saw in videos of mathematics lessons obtained in a pilot study. Despite these methodological differences and challenges, the data collected with the self-report questionnaire and the observation instrument revealed a similar pattern of relatively frequently and infrequently used strategies for differentiation and thus corroborated each other. Moreover, the complementary use of both instruments exploited the strengths of both types of measures: The self-report questionnaire was administered in large sample and included aspects of differentiation which are hard to observe (e.g., the analysis of students' written work), whereas the video observations were used to examine differentiation in more detail in a minute-by-minute fashion.

An implication of the substantial differences between the level of use of various strategies for differentiation is that evaluating the level of use of specific strategies may be more meaningful than evaluating the overall level of differentiation. At least, researchers should be aware that general, single questions such as 'how often do you differentiate in your classroom?' (Dixon, Yssel, McConnell, & Hardin, 2014, p. 119) may yield data which are hard to interpret. The assessment of specific strategies for differentiation is likely to be more informative for both research, policy and practice since this provides more directions for areas in need of improvement. The findings of the current study point to the following specific areas for improving the implementation of differentiation. First, the use of specialised adaptations that are carefully matched to students' diagnosed educational needs could be improved. Second, teachers could give more attention to high-achieving students, especially in adapted instruction.

6.3 Goal N° 3: Effects of professional development (PD)

Based on the specification of differentiation, a yearlong PD programme about the cycle of differentiation and the belonging strategies for differentiation in primary school mathematics was developed and implemented in collaboration with the consortium of experts. Schools in Cohort 1 received the PD in Year 1, schools in Cohort 2 received the PD in Year 2, and schools in Cohort 3 were in a control condition for two years. The effects of the programme on teachers' instructional behaviour as well as on students' achievement were evaluated. The results reported in Chapter 4 provide partial support for positive effects on the most distal outcome, i.e., student achievement on a standardised mathematics test. In Year 1, students of Cohort 1 – whose teachers participated in the PD programme in that year – demonstrated more achievement growth than students of the other cohorts. This applied to students of all achievement levels. These results were controlled for grade level, nonverbal intelligence, visual-spatial and verbal working memory (and teased apart from potential differences in the level of achievement due to the use of latent growth models in which the level of achievement is distinguished from the rate of achievement growth). Moreover, schools were randomly assigned to cohorts and seemed to be largely comparable at baseline in terms of student achievement (after controlling for grade level), nonverbal intelligence, and working memory, as well as years of experience at the teacher level. Furthermore, the results of the evaluation questionnaire indicated that teachers experienced the PD programme as useful and reported to use what they learned in their daily practice.

Despite these indications that the PD programme had a positive effect on student achievement growth in Year 1, the results of the video observations reported in Chapter 3 could not demonstrate unequivocally that the PD programme had a positive effect on teachers' observable implementation of differentiation. Teachers did not implement more differentiation directly at the end of the year in which they participated in the intervention. Teachers of Cohort 1 did implement more differentiation than teachers of the other cohorts in Year 2 (i.e., the year after Cohort 1 participated in the PD), which might be interpreted as a long-term effect of the intervention. However, alternative explanations could not be excluded since the between-cohort difference seemed to be partly due to a decrease of observed implementation in Cohort 2 and 3 rather than due to an increase in Cohort 1. The following limitations may explain why it was hard to demonstrate observable effects on teachers' implementation of differentiation. First, detecting observable changes in teachers' behaviour is challenging in general. For the well-established Mathematical Quality of Instruction instrument (Learning Mathematics for Teaching Project, 2011), the use of three observations per teacher – each scored by two observers – was recommended based on a generalisability study to attain sufficient reliability for research purposes (Hill, Charalambous & Kraft, 2012). Due to resource limitations, observing each teacher twice per timepoint by a single observer was the maximum possible in our study. It is unknown whether the resulting data were sufficiently reliable to measure changes in teachers' observable behaviour over time, especially since a new observation instrument was used. Second, any changes in teachers' behaviour may have been rather subtle and therefore hard to measure. For example, most teachers already worked with ability groups at baseline. During the PD programme, attention was spent on how the groups should be created (i.e., based on which measures, how to combine information from various sources, how to identify the educational needs of the students in the groups, the importance of flexibility of the subgroups). This may have resulted in a more refined grouping of students based on a more careful diagnosis of educational needs, which may have had positive effects on achievement, but which would have been hard to observe during a mathematics lesson. Specifically, the used observation instrument does not measure the process of identification of educational needs and only focuses on observable adaptations of instruction and practice during the observed lesson. Third, the video-observations were carried out in a relatively small subsample of teachers due to the highly intensive coding process. Taken together, the design of this research project may have been more sensitive for detecting changes in student achievement than for detecting changes in teachers' instructional behaviour due to the potentially higher reliability and validity of the standardised mathematics achievement test as well as the larger sample size of students.

Another unexpected result was that the positive achievement effects could not be replicated in Cohort 2 in Year 2. As has been discussed in Chapter 4, this might be explained by practical factors: First, schools in Cohort 2 had to wait for one year before they could start the PD programme. This may have affected their motivation negatively, since schools may have identified other priorities for PD in the meantime. Second, in Year 2, Cohort 2 included relatively many teachers who were new at the school or who had fewer years of teaching experience compared to teachers of the other cohorts. Due to these circumstances, teachers may have needed to spend their attention on more basic aspects of teaching (e.g., classroom management) rather than on differentiation. Third, relatively many schools in Cohort 2 started to use a new mathematics curriculum over the course of the study. This may have drawn teachers' attention towards implementation of the new curriculum rather than to the implementation of differentiation in mathematics. In this study, we tried to enhance the practical validity of the findings by implementing the PD programme in a real-life context. However, this diminished the experimental control. When planning research in which a delayed intervention condition is an option, researchers should consider whether and how the delay itself might affect the effects of the intervention. In the dynamic context of primary schools, a delayed intervention condition may not be the best option.

Compared to other PD programmes about differentiation (e.g., Johnsen, Haensly, Ryser, & Ford, 2002; Klingner, Ahwee, Pilonieta, & Menendez, 2003; VanTassel-Baska et al., 2008), the current programme attempted to cover relatively much content in relatively little time. The broadness of the programme was a deliberate choice – we wanted to promote differentiation as a way to adapt education to the diverse educational needs of students with a broad range of achievement levels, rather than focusing on one group of students (with the risk of fulfilling the needs of one group of students at the expense of another group). Moreover, the cycle of differentiation resulting from the specification phase clearly showed that differentiation is a complex process affecting all phases of teaching, from lesson preparation through evaluation. Given that the steps of the cycle of differentiation are interdependent (i.e., an appropriate adaptation of instruction should be based upon adequately diagnosed educational needs and appropriate goals), all steps needed to be covered in the PD programme. On the other hand, the risk of a programme that is too broad is that its content is not covered in sufficient depth. In the current programme, schools were required to cover all steps of the cycle of differentiation within one year – although they also had some freedom to zoom in on topics which were particularly relevant for that school. Previous studies have indicated that it may take more time (two to three years of intensive PD) to produce substantial changes in teachers' behaviour (Borko, Mayfield, Marion, Flexer, & Cumbo, 1997; VanTassel-Baska et al., 2008). A potential solution could be to use the cycle of differentiation as an overarching framework for PD during a longer period of time. After a general introduction of the cycle of differentiation, schools could cover particular aspects of differentiation in more depth each year, for example spending attention on subgroup instruction for high-achieving students in one year and learning how to use diagnostic conversations in another year. However, it is the question whether such extended PD would be realistic in the Dutch primary school context.

Given these complex findings and limitations, what can we conclude regarding the effects of the PD programme and are the implications for educational practice? The positive achievement effects in Cohort 1 indicate that students of all achievement levels may benefit from a teacher-level intervention in which teachers learn how to (better) implement differentiation in mathematics. Although the effects of PD in Year 1 on the slope of achievement growth (with standardised effects between 0.10 and 0.15) were relatively small, they were similar to or larger than the effects of nonverbal intelligence and working memory on the slope of achievement growth. Moreover, the effects were similar in size and direction for low-achieving, average-achieving and high-achieving students, indicating that students of all achievement levels profited from the intervention about equally. Thus, PD about differentiation did not have adverse effects for low-achieving students, in contrast to some previous studies in which negative effects of ability grouping were found for low-achieving students (reviewed by Deunk, Doolaard, Smale-Jacobse, & Bosker, 2015). This may be explained by the fact that we did not evaluate the effect of ability grouping *per se*, but rather the effects of PD which was intended to enable teachers to make better adaptations to the needs of both low-achieving, average-achieving, and high-achieving students. However, the positive achievement effects for students of all achievement levels in Cohort 1 could not be replicated in Cohort 2 and positive effects on teachers' observable implementation of differentiation could not be unequivocally demonstrated. Taken together, our findings imply that PD about differentiation in primary mathematics education has the potential to raise student achievement, but such effects are not guaranteed. Given that most teachers already implement some aspects of differentiation before PD about this topic, the differences between teachers' implementation of differentiation before and after PD may be rather subtle – although subtle differences in how differentiation is implemented may make a substantial qualitative difference in meeting students' educational needs. Making these subtle qualitative changes seems to require relatively long-term and intensive PD, but this may be worthwhile since it can have small but meaningful positive effects on the achievement of all students.

6.4 Goal N° 4: Relations between achievement and motivation for mathematics

Chapter 5 examined the relations between achievement and motivation for mathematics, with particular attention for potential differences between low-achieving, average-achieving, and high-achieving students. Specifically, we examined whether several core aspects of motivation for mathematics – self-efficacy, self-concept, task value, and mathematics anxiety – were predicted by previous achievement in mathematics. Moreover, we examined how these motivational variables jointly predicted subsequent mathematics achievement and whether they mediated between previous and subsequent mathematics achievement. These relations were first investigated in the total sample and subsequently in subsamples of low-achieving, average-achieving and high-achieving students. We reasoned that these relations might differ between achievement groups since the educational experiences of low-achieving, average-achieving, and high-achieving students differ substantially.

In the total sample, motivation was shown to be substantially related to previous achievement. Previous achievement had positive effects on perceived competence and task value and negative effects on mathematics anxiety. Moreover, all motivational variables were interrelated (the relation between self-efficacy and self-concept was so strong that these two variables needed to be combined into a single perceived competence variable to circumvent multicollinearity problems). Thus, students with relatively low scores on previous mathematics tests – who may have experienced little success in mathematics, possibly for an extended period of time – are unlikely to feel competent in mathematics or to enjoy mathematics whereas they are more likely to be anxious for mathematics. Regarding the effects of motivation on subsequent achievement, perceived competence was the only motivational variable which significantly predicted subsequent achievement after controlling for previous achievement (yielding a small but significant indirect effect from previous achievement on subsequent achievement through perceived competence). Thus, even if students score equally on previous achievement and working memory, students who feel more competent are likely to achieve higher in the future. In contrast, task value and mathematics anxiety did not predict subsequent achievement, which might be explained by the inclusion of previous achievement as well as multiple motivational variables in the model (due to their shared variance; see Chapter 5).

The analysis of subsamples of low-achieving, average-achieving, and high-achieving students demonstrated that these relations were largely similar across achievement groups. However, there were some variations in the effects of previous achievement on motivation, which were generally strongest in the high-achieving group. We speculated that an evaluation of one's own achievement relative to students of a similar achievement level (i.e., within the subsample) might be more salient for high-achieving students than for low-achieving students (e.g., it could be that low-achieving students primarily compare their own achievement to the achievement of the whole class rather than to other low-achieving students, whereas high-achieving students might compare themselves more often to other high-achieving students). An alternative possibility is that high-achieving students might be better able to construct realistic perceptions of their own abilities based on previous achievement. Our study does not inform about these processes, but these are interesting hypotheses to explore in future research.

The effects of motivation on subsequent achievement were similar across achievement groups. Mathematics anxiety and task value did not have a significant effect on subsequent achievement in any of the achievement groups. Thus, it is not the case that these aspects of motivation are only related to subsequent achievement in one particular achievement group (which might have been a potential explanation for the lack of effects in the total sample). In all achievement groups, perceived competence had a small but significant effect on subsequent achievement. Thus, the relative importance of diverse aspects of motivation seems to be similar across achievement groups. One of the reasons for carrying out this study in the context of project GROW was that knowledge about potential differences regarding these relations between achievement groups might provide directions for differentiating instruction based on potentially varying motivational needs across achievement groups. Our results provide no indications that specific aspects of motivation should be differentially fostered depending on students' achievement level. Rather, the positive effect of perceived competence across achievement groups implies that fostering perceived competence might be a promising way to enhance achievement for students of all achievement levels. However, this conclusion should be drawn with care since our results also indicate that actual achievement is an important source of perceived competence. It does not seem to make sense to foster unrealistically high perceived competence, but it might be helpful to create situations in which students can experience mastery. This might enhance students' perceived competence and, in turn, adaptive learning behaviours and subsequent achievement. Experimental studies have provided initial indications that perceived competence can be manipulated and that experiencing success may have positive effects on productive learning behaviour and achievement (e.g., Jansen et al., 2013; Sheldon & Filak, 2008).

A strength of this study is that the effects of motivation on achievement were controlled for previous achievement as well as working memory. A limitation is that the effects of achievement on motivation were not controlled for previous motivation. This might explain why the effects of achievement on motivation were generally stronger in this study than the effects of motivation on achievement. In future research, cross-lagged designs in which both motivation and achievement are measured repeatedly over multiple timepoints could be used to examine the interplay between motivation and achievement in more detail. On the other hand, the temporal ordering of the measures – with achievement measured at T1 and T3 and motivation measured at T2 – was very suitable to test mediation.

6.5 General reflections and directions for future research

6.5.1 Tension between desirability and feasibility of differentiation

Throughout this research project, we repeatedly encountered a tension between what would be desirable from the perspective of meeting students' educational needs versus what was realistic and feasible in teachers' daily practice. Even if teachers have the necessary knowledge and skills for implementing the recommended strategies for differentiation, they also need to find the time to do it: both during the lesson (e.g., diagnostic conversations with individual students or subgroup instruction for high-achieving students) as well as before and after the lesson (e.g., preparing a broad whole-class instruction which is relevant for students of diverse achievement levels or analysing students' work to diagnose educational needs). Project GROW was executed from a professional development perspective and therefore focused on teachers' adaptive teaching competence. However, it should be kept in mind that the implementation of differentiation is influenced by other factors too. Contextual factors such as the time allotted to teachers for preparing their lessons may either help or hinder the implementation of differentiation (Puzio, Newcomer, & Goff, 2015; Roiha, 2014). Policymakers should be aware that the demands currently placed on teachers in Dutch primary education are very high – especially given that mathematics is only one of many subjects and that students' current achievement level is only one of many potential grounds for differentiation (besides other sources of diversity such as developmental disorders or migration backgrounds). The transition towards inclusive education has not been coupled with a substantial increase of supports for general education teachers in the Netherlands. Policymakers and school administrators should consider how they can support teachers in implementing differentiation, for example by hiring more teaching assistants or by giving teachers more preparation time. Researchers could help policy makers to make informed decisions by providing more data on which of these provisions are most supportive for teachers (e.g., should money preferably be invested in more teaching assistants or in smaller classes?).

Another direction for future research is the investigation of other, potentially less expensive, ways to support teachers in the implementation of differentiation, such as pre-differentiated curricula or technological applications for differentiation. The tension between desirability and feasibility becomes visible again in these types of supports for differentiation. In the current study, we saw that many teachers used the suggestions for differentiation as provided by the curriculum. On the one hand, pre-differentiated curricula may help teachers to implement differentiation without increasing their workload too much. On the other hand, too much pre-designed differentiation may also have adverse effects. Some curricula provide separate workbooks for low-achieving students from grade 1 onwards. The risk of such an approach is that both teachers and students may be satisfied when a low-achieving student successfully completes the assignments in a '1-star workbook'. Using separate workbooks may also decrease the flexibility of the subgroups, since this may make it less convenient for low-achieving students to attempt the regular tasks if appropriate in a specific lesson or content area. Such processes may reduce the chance that low-achieving students can catch up with typically-achieving peers over time. More generally, pre-differentiated curricula may reduce teachers' own initiative for evaluating critically what students need and making appropriate adaptations. For example, teachers may simply assign the pre-designed enrichment tasks to high-achieving students without considering whether these tasks are suitable for meeting the educational needs of these individual students. Future research should investigate how suggestions for differentiation can be made in such a way that teachers can use them as necessary, but maintain their responsibility for diagnosing students' educational needs and making adaptations in response to these.

The issue of feasibility applied not only to the implementation of differentiation but also to the PD programme itself. A yearlong PD programme consisting of ten team meetings coupled with additional reading and preparation represents a substantial increase of teachers' already high workload. Teachers therefore perceived the programme as rather intensive. From a researcher perspective, an even longer and perhaps also more intensive PD programme might be desirable to cover the underlying knowledge and skills required for implementing the recommended strategies for differentiation in sufficient depth. For example, the refined diagnosis of students' educational needs requires much underlying knowledge of how students learn mathematics (e.g., about the order in which students typically learn mathematical concepts, recommended solution strategies, and common misconceptions). Building such knowledge takes time. While PD may help to extend such knowledge, it also seems to be important that initial teacher education equips beginning teachers with a solid base of didactical knowledge in the domain of mathematics. Acquiring a systematic knowledge base is likely to be more feasible during initial teacher education than during the in-service teaching years (i.e., besides a busy teaching job). In-service teachers could use this knowledge base as a foundation for building adaptive teaching competency through increased teaching experience and professional development.

6.5.2 Observing differentiation

A new observation instrument, the Differentiation in Mathematical Instruction (DMI), was developed to measure the implementation of differentiation in mathematics for students of diverse achievement levels (i.e., both low-achieving and high-achieving) in more detail than previous studies. However, observing differentiation turned out to be a complex issue which requires more attention in future research. It was challenging to strike a good balance between reliability on the one hand and validity on the other hand. Providing standardised low-inference scoring guidelines may enhance interrater reliability but can make it hard to measure the quality of differentiation, since differentiation is a complex process in which subtle differences regarding *how* an adaptation is used can make a large qualitative difference. These qualitative differences, however, are hard to observe since the observer often does not have sufficient information to determine whether a particular adaptation is indeed suitable for meeting the educational needs of a particular student. Students' educational needs are not always directly visible on a video of a mathematics lesson, and may (or may not) have been determined by the teacher previously based on other sources of information. To circumvent these problems, the DMI observation instrument focused on the use of observable adaptations recommended for (groups of) low-achieving and high-achieving students in general, but the instrument did not capture the match between educational needs and selected adaptations (e.g., whether the adaptations made by the teachers are indeed appropriate for individual students). While this would be very complicated to measure, it should be attempted in future research because this match is important for the quality of differentiation. Using multiple methods (e.g., extending a lesson observation with a teacher interview in which the teacher can explain her decisions, or collecting students' written work) might be helpful. An alternative and perhaps more feasible way could be to extend the DMI with items on *whether* the teacher engages in observable diagnostic activities during the lesson prior to making adaptations. Although this would still not capture the quality of the match between educational needs and instructional adaptations, it would provide more information regarding the extent to which teachers engage in diagnostic activities during the mathematics lesson. Based on what we saw (or did not see) in the videos collected for this project, it seems that teachers' attention for determining students' educational needs before starting to make adaptations during instruction to subgroups or individual students is limited.

6.5.3 Effects of differentiation on motivation

The results regarding the relations between achievement and motivation – in particular, the demonstrated importance of perceived competence – raise interesting questions concerning the effects of differentiated instruction using within-class ability groups on both perceived and actual mathematical competence. Some previous studies have found negative effects of ability grouping for students placed in low-ability groups (reviewed by Deunk et al., 2015). Ability grouping might make students more aware of their own achievement relative to their peers. If placement in a low-achieving subgroup reduces students' perceived competence for mathematics, it might function as a self-fulfilling prophecy. Suggestions such as ensuring that the subgroups are flexible and promoting a dynamic rather than fixed view of students' abilities have been made (e.g., Tomlinson et al., 2003), but more research is necessary to establish whether these conditions are sufficient to overcome potential negative effects of placement in a low-ability group on perceived competence. On the other hand, homogeneous within-class ability grouping might also have positive effects on perceived competence, since it gives students the opportunity to interact more with students of a similar achievement level. Thus, low-achieving students might start to compare their own achievement to other low-achieving students rather than to the class-average, which might reduce the big-fish-little-pond effect (referring to the fact that students evaluate their own achievement relative to the achievement of other students; students with a score of 6/10 are likely to have a higher self-concept in a class in which the average score is 6/10 than in a class in which the average score is 8/10; Marsh, 1987). Indeed, Roy, Guay, and Valois (2015) found that the big-fish-little-pond effect was less pronounced for low-achieving students if teachers frequently used differentiated instruction. Moreover, when instruction and practice are better adapted to the needs of the students due to the use of differentiated instruction, especially low-achieving students might experience less failure and more success – which might have positive effects on perceived competence (Usher & Pajares, 2008). Furthermore, if differentiated instruction has positive effects on students' actual achievement, this may have positive effects on students' perceived competence in turn (Marsh & Martin, 2011). In sum, the effects of differentiated instruction using within-class ability groups on students' perceived and actual mathematical competence are likely to be complex – possibly with simultaneous positive and negative effects due to different underlying mechanisms - and more research is necessary to unravel these processes.

6.6 General conclusions and implications

Taking all studies in this dissertation together, the main conclusions and implications of project GROW can be summarised as follows. Differentiation is a complex and multifaceted construct. It may be organised using within-class achievement groups, but it involves much more than that: diagnosing educational needs, setting differentiated goals, adapting instruction, adapting practice, and evaluating students' learning progress and process. The cycle of differentiation as well as the recommended strategies for each step provide teachers and teacher educators with concrete guidelines for implementing differentiation in primary mathematics education.

Regarding the implementation of differentiation, it was found that most teachers implement at least some aspects of differentiation, but substantial differences between the level of use of various strategies for differentiation were revealed. The results of this dissertation provide specific directions for improving the implementation of differentiation. First, the refined diagnosis of students' educational needs as well as specialised adaptations to meet these needs could be improved. Second, the specific needs of high-achieving students deserve more attention, especially during (subgroup) instruction.

The results of the PD programme evaluated in this dissertation show that PD about differentiation in primary mathematics education has the potential to improve student achievement growth in mathematics. Importantly, students of all achievement levels can profit from teacher-level PD about differentiation. However, such effects are not guaranteed, indicating that enhancing differentiation is not straightforward (and neither is its measurement).

Regarding the relations between motivation and achievement, it was found that the effects of previous achievement were generally stronger for high-achieving students, whereas the effects of motivation on subsequent achievement were similar across achievement groups. For students of all achievement levels, perceived competence was the only motivational variable which was not only influenced by previous achievement, but also predicted subsequent achievement after controlling for previous achievement and working memory. These findings imply that fostering students' perceived competence might be a potential way to enhance student achievement. At the same time, the bidirectional relation between perceived competence and achievement raises important questions for future research regarding the effects of differentiated instruction on the perceived competence of students of diverse achievement levels.

Given the complexity of implementing differentiation, in addition to the other demands placed on teachers, teachers should be offered sufficient support. The question of how teachers can be supported most effectively deserves more attention in future research, but policymakers should also be aware that successful inclusive education does not come for free. Teachers, teacher educators, school administrators, policymakers and researchers should work together to achieve a shared goal: to differentiate education in such a way that all students are given the chance to develop their potential.

References

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education*, *13*, 259-278.

Brighton, C. M., Moon, T. R., & Huang, F. H. L. (2015). Advanced readers in reading first classrooms: Who was really "Left behind"? Considerations for the field of gifted education. *Journal for the Education of the Gifted*, *38*, 257-293. doi:10.1177/0162353215592501

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*, 181-200. doi:10.3102/0013189X08331140

Deunk, M., Doolaard, S., Smale-Jacobse, A., & Bosker, R. J. (2015). *Differentiation within and across classrooms: A systematic review of studies into the cognitive effects of differentiation practices*. Groningen, The Netherlands: GION.

Dixon, F. A., Yssel, N., McConnell, J. M., & Hardin, T. (2014). Differentiated instruction, professional development, and teacher efficacy. *Journal for the Education of the Gifted*, *37*, 111-127. doi:10.1177/0162353214529042

Hertberg-Davis, H. (2009). Myth 7: Differentiation in the regular classroom is equivalent to gifted programs and is sufficient: Classroom teachers have the time, the skill, and the will to differentiate adequately. *Gifted Child Quarterly*, *53*, 251-253. doi:10.1177/001698620934692

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, *41*, 56-64. doi: 10.3102/0013189X12437203

Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H. G., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, *24*, 190-197. doi:10.1016/j.lindif.2012.12.014

Johnsen, S. K., Haensly, P. A., Ryser, G. A. R., & Ford, R. F. (2002). Changing general education classroom practices to adapt for gifted students. *Gifted Child Quarterly*, *46*, 45-63. doi: 10.1177/001698620204600105

Klingner, J. K., Ahwee, S., Pilonieta, P., & Menendez, R. (2003). Barriers and facilitators in scaling up research-based practices. *Exceptional Children*, *69*, 411-429. doi:10.1177/001440290306900402

KNAW. (2009). *Rekenonderwijs op de basisschool: Analyse en sleutels tot verbetering* [Mathematics education in primary school: Analysis and keys to improvement]. Amsterdam, The Netherlands: KNAW.

Learning Mathematics for Teaching Project. (2011). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education*, *14*, 25-47. doi:10.1007/s10857-010-9140-1

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*, 280-295.

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*, 59-77. doi:10.1348/000709910X503501

McLeskey, J., & Waldron, N. L. (2011). Educational programs for elementary students with learning disabilities: Can they be both effective and inclusive? *Learning Disabilities Research & Practice*, *26*, 48-57. doi:10.1111/j.1540-5826.2010.00324.x

Meelissen, M. R. M., & Punter, R. A. (2016). *Twintig jaar TIMSS: Ontwikkelingen in leerlingprestaties in de exacte vakken in het basisonderwijs 1995 - 2015* [Twenty years of TIMSS: Developments in student achievement in mathematics and science in primary education 1995 - 2015]. Enschede, The Netherlands: Universiteit Twente.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System: Manual K-3*. Baltimore, MD: Paul H. Brookes.

Puzio, K., Newcomer, S. N., & Goff, P. (2015). Supporting literacy differentiation: The principal's role in a community of practice. *Literacy Research and Instruction*, *54*, 135-162. doi: 10.1080/19388071.2014.997944

Roiha, A. L. (2014). Teachers' views on differentiation in content and language integrated learning (CLIL): Perceptions, practices and challenges. *Language and Education*, *28*, 1-18. doi: 10.1080/09500782.2012.748061

Roy, A., Guay, F., & Valois, P. (2013). Teaching to address diverse learning needs: Development and validation of a differentiated instruction scale. *International Journal of Inclusive Education*, *17*, 1186-1204. doi:10.1080/13603116.2012.743604

Roy, A., Guay, F., & Valois, P. (2015). The big-fish–little-pond effect on academic self-concept: The moderating role of differentiated instruction and individual achievement. *Learning and Individual Differences*, *42*, 110-116. doi:10.1016/j.lindif.2015.07.009

Rubenstein, L. D., Gilson, C. M., Bruce-Davis, M. N., & Gubbins, E. J. (2015). Teachers' reactions to pre-differentiated and enriched mathematics curricula. *Journal for the Education of the Gifted*, *38*, 141-168. doi:10.1177/0162353215578280

Sheldon, K. M., & Filak, V. (2008). Manipulating autonomy, competence and relatedness support in a game-learning context: New evidence that all three needs matter. *British Journal of Social Psychology*, *47*, 267-283.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., ... Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, *27*, 119-145. doi:10.1177/016235320302700203

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, *78*, 751-796. doi:10.3102/0034654308321456

Van de Weijer-Bergsma, E., Van Luit, J. E. H., Prast, E. J., Kroesbergen, E. H., Kaskens, J., Compagnie-Rietberg, C., ... Logtenberg, H. (2016). *Differentiëren in het rekenonderwijs: Hoe doe je dat in de praktijk?* [Differentiation in mathematics education: How to do that in practice?]. Doetinchem, The Netherlands: Graviant scientific & educational books.

VanTassel-Baska, J., Feng, A. X., Brown, E., Bracken, B., Stambaugh, T., French, H., ... Bai, W. (2008). A study of differentiated instructional change over 3 years. *Gifted Child Quarterly*, *52*, 297-312. doi:10.1177/0016986208321809

VanTassel-Baska, J., & Stambaugh, T. (2005). Challenges and possibilities for serving gifted learners in the regular classroom. *Theory into Practice*, *44*, 211-217. doi:10.1207/s15430421tip4403_5

Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, *25*, 1051-1060. doi:10.1016/j.tate.2009.04.002


MEES
Teken lijnen.
Teken lijnen in je schrift. Meet ze.
Schrijf erbij hoe lang ze zijn.
Hoe lang? Gebruik je liniaal.

Nederlandse samenvatting
Summary in Dutch

Achtergrond

Iedere leerkracht krijgt te maken met verschillen tussen leerlingen in de klas. Leerlingen verschillen van elkaar op allerlei gebieden zoals intelligentie, werkgeheugen, motivatie, sociaaleconomische status en eerdere leerervaringen. Deze factoren zijn onder andere van invloed op de leerprestaties en op wat leerlingen nodig hebben om een stapje verder te komen in hun leerproces. Onderwijsbehoeften, oftewel wat leerlingen nodig hebben om een bepaald onderwijsdoel te bereiken, variëren dus tussen leerlingen. In het basisonderwijs zijn deze verschillen bijzonder uitgesproken, omdat leerlingen van alle leerniveaus bij elkaar in de klas zitten (in tegenstelling tot op de middelbare school, waar in Nederland een onderscheid wordt gemaakt tussen vwo, havo en vmbo, en daarbinnen diverse uitstroomprofielen). Bovendien wordt de diversiteit aan onderwijsbehoeften steeds groter doordat leerlingen met speciale onderwijsbehoeften, die vroeger naar het speciaal basisonderwijs zouden zijn gegaan, tegenwoordig zo veel mogelijk in reguliere basisschoolklassen worden geplaatst (inclusie, oftewel Passend Onderwijs). Voor de leerkracht kan het een hele uitdaging zijn om tegemoet te komen aan deze verschillende onderwijsbehoeften van de leerlingen in de klas.

Het proces van afstemmen op verschillende onderwijsbehoeften noemen we differentiatie. Dit proefschrift richt zich op differentiatie op basis van het huidige leerniveau, ook wel niveaudifferentiatie genoemd. Niveaudifferentiatie omvat twee belangrijke componenten: het analyseren van het huidige leerniveau om daarmee de onderwijsbehoeften vast te stellen en het maken van aanpassingen in doelen, instructie en verwerking om de vastgestelde onderwijsbehoeften te vervullen.

Het toepassen van niveaudifferentiatie is sterk verbonden met een bepaalde vakinhoud: om te differentiëren in het vak rekenen moeten leerkrachten bijvoorbeeld veel vakdidactische kennis hebben over de ontwikkeling van rekenvaardigheden, rekenstrategieën, en mogelijke aanpassingen in de rekeninstructie bij specifieke onderwijsbehoeften in dit vak. In dit proefschrift staat daarom één vak centraal, namelijk rekenen. Het vak rekenen staat in de belangstelling, omdat er zorgen zijn over het rekenniveau van Nederlandse leerlingen in vergelijking met leerlingen uit andere landen. Nascholing van leerkrachten over rekenonderwijs in het algemeen, en in het bijzonder over differentiatie in het rekenonderwijs, wordt gezien als een mogelijke manier om de rekenprestaties van Nederlandse leerlingen te verbeteren.

Project Gedifferentieerd RekenOnderWijs

Tegen deze achtergrond is het project Gedifferentieerd RekenOnderWijs (GROW) geïnitieerd om differentiatie in het rekenonderwijs te bestuderen en te verbeteren. In het project GROW hebben onderzoekers van de Universiteit Utrecht nauw samengewerkt met een consortium van leerkrachtopleiders en onderwijsadviseurs met expertise op het gebied van rekenonderwijs en differentiatie. Deze praktijkexperts vormden de schakel tussen onderwijsonderzoek en onderwijspraktijk. In de eerste fase van het project (schooljaar 2011–2012) is consensus gezocht tussen de consortiumleden over hoe differentiatie in het rekenonderwijs eruit zou moeten zien. Op basis van de resultaten is een nascholingstraject ontwikkeld over differentiatie in het rekenonderwijs. In de tweede fase van het project (schooljaren 2012–2013 en 2013–2014) zijn de effecten van dit nascholingstraject geëvalueerd in een grootschalige studie. Hiertoe werden de 32 deelnemende basisscholen random verdeeld over drie cohorten. Scholen uit Cohort 1 namen deel aan het nascholingstraject in Jaar 1, scholen uit Cohort 2 namen deel aan het nascholingstraject in Jaar 2 en scholen uit Cohort 3 vormden twee jaar lang een controleconditie. Op alle scholen zijn gedurende twee jaar data verzameld op leerkrachtniveau (toegepaste differentiatie) en op leerlingniveau (rekenprestaties, motivatie, werkgeheugen, en non-verbale intelligentie).

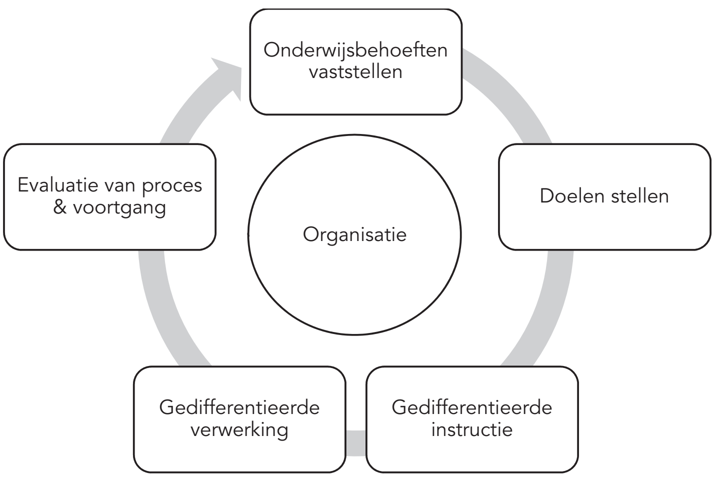
Doelen van dit proefschrift

Voor dit proefschrift zijn de volgende vier doelen gesteld, elk gerelateerd aan het centrale thema ‘differentiatie in het rekenonderwijs’. Het eerste doel was om nader te specificeren wat differentiatie in het rekenonderwijs inhoudt volgens experts. Het tweede doel was om te onderzoeken in hoeverre verschillende aanbevolen strategieën voor differentiatie al worden geïmplementeerd (zonder specifieke nascholing). Het derde doel was om de effecten van het ontwikkelde nascholingstraject over differentiatie in het rekenonderwijs te evalueren. Hierbij zijn zowel de effecten op de implementatie van differentiatie door leerkrachten als de effecten op de rekenprestaties van de leerlingen onderzocht. Het vierde doel was om de relaties tussen rekenprestaties en verschillende aspecten van motivatie voor rekenen te onderzoeken met speciale aandacht voor mogelijke verschillen tussen laagpresterende, gemiddeld presterende en hoogpresterende leerlingen. In deze samenvatting worden eerst de belangrijkste resultaten per doel besproken, en daarna de algemene conclusies en implicaties.

Doel 1: Specificatie van differentiatie in het rekenonderwijs

In hoofdstuk 2 is gespecificeerd wat differentiatie inhoudt in de context van rekenonderwijs op de basisschool. Door middel van focusgroepdiscussies en vragenlijsten hebben de experts in het consortium consensus bereikt over de vraag hoe goede differentiatie in het rekenonderwijs eruit zou moeten zien. Dit heeft het volgende model voor differentiatie opgeleverd.

De differentiatiecyclus, afgebeeld in Figuur 1, begint met het vaststellen van de onderwijsbehoeften van de verschillende leerlingen in de klas (stap 1). Hierbij worden zowel formele als informele informatiebronnen gebruikt (respectievelijk bijvoorbeeld rekentoetsen en diagnostische gesprekken). In eerste instantie worden de leerlingen ingedeeld in drie niveaugroepen op basis van hun scores op formele rekentoetsen. In de loop van het schooljaar wordt deze indeling verfijnd en waar nodig bijgesteld op basis van aanvullende informatie zoals het dagelijks rekenwerk van de leerlingen. De niveaugroepen zijn flexibel: leerlingen kunnen wisselen tussen de groepen op basis van veranderingen in hun onderwijsbehoeften. In stap 2 worden gedifferentieerde doelen gesteld die uitdagend en realistisch moeten zijn voor de verschillende (groepen) leerlingen. In stap 3 wordt de instructie gedifferentieerd. De klassikale instructie wordt breed opgezet om deze geschikt te maken voor leerlingen met verschillende leerniveaus. In subgroepinstructie kan worden ingespeeld op de gedeelde onderwijsbehoeften van kinderen met een vergelijkbaar leerniveau. Zowel laagpresterende als hoogpresterende leerlingen hebben behoefte aan specifieke instructie en begeleiding. Tot slot kan de leerkracht individuele instructie inzetten om tegemoet te komen aan specifieke onderwijsbehoeften van individuele leerlingen. In stap 4 wordt de verwerking gedifferentieerd. Hierbij kan gedacht worden aan de inhoud en hoeveelheid van de verwerkingsopdrachten, maar ook aan de manier waarop verwerkingsopdrachten gemaakt worden (bijvoorbeeld met of zonder gebruik van ondersteunende materialen). In stap 5 wordt de voortgang en het leerproces van de leerlingen geëvalueerd. Hierbij wordt niet alleen gekeken of de leerdoelen behaald zijn, maar ook of de gekozen aanpassingen van instructie en verwerking effectief zijn gebleken



Figuur 1 De differentiatiecyclus.

voor de leerlingen. Dit geeft weer nieuwe informatie over de onderwijsbehoeften van de leerlingen in de klas en daarmee is de differentiatiecyclus rond. Organisatie staat in het midden van de differentiatiecyclus, omdat organisatorische aspecten aandacht vragen bij iedere stap. Hierbij kan bijvoorbeeld gedacht worden aan effectief klassenmanagement voor het werken met subgroepen en aan langetermijnplanning voor differentiatie. Voor iedere stap van de cyclus zijn diverse aanbevolen strategieën voor differentiatie gespecificeerd door de experts in het consortium.

Doel 2: Implementatie van differentiatie

In hoofdstuk 2 en 3 is onderzocht in hoeverre leerkrachten de verschillende aanbevolen strategieën voor differentiatie al toepasten in hun rekenonderwijs (vóór specifieke nascholing over differentiatie). Om dit te kunnen meten, zijn twee nieuwe instrumenten ontwikkeld: een zelfinschattingsvragenlijst en een video-observatieinstrument. De zelfinschattingsvragenlijst is ingevuld door 268 leerkrachten aan het begin van het onderzoek (hoofdstuk 2). Bij een deelsteekproef van 55 leerkrachten zijn daarnaast video-opnames gemaakt van twee rekenlessen, die vervolgens zijn gescoord met het video-observatieinstrument (hoofdstuk 3). Deze twee informatiebronnen hebben de volgende bevindingen opgeleverd.

Veel leerkrachten werkten al met homogene subgroepen op basis van rekenniveau. Uit de specificatie van differentiatie (zie doel 1) was echter gebleken dat differentiatie méér inhoudt dan het werken met niveaugroepen: het is vooral belangrijk dat de niveaugroepen worden gebruikt om instructie en verwerking aan te passen aan de onderwijsbehoeften van de leerlingen in de subgroepen. Hierbij werd duidelijk een verschillende benadering gekozen voor laagpresterende versus hoogpresterende leerlingen. Voor laagpresterende leerlingen lag de nadruk op het differentiëren van instructie, bijvoorbeeld door middel van verlengde instructie, in een subgroep na afloop van de klassikale instructie. Het aanpassen van de verwerkingstaken kreeg minder nadruk maar kwam ook voor, met name als de rekenmethode aangepaste verwerkingsopdrachten voor laagpresterende leerlingen aanbood. Voor hoogpresterende leerlingen, daarentegen, was het patroon omgekeerd. De nadruk lag op het differentiëren van de verwerkingsopdrachten, met name met behulp van verrijkingstaken die in de rekenmethode standaard waren opgenomen. Specifieke instructie of begeleiding voor hoogpresterende leerlingen werd echter nauwelijks geobserveerd. Dit is een punt voor verbetering omdat ook hoogpresterende leerlingen behoefte hebben aan specifieke instructie of begeleiding, zeker als zij werken aan uitdagende verrijkingstaken.

Verder werden algemene differentiatiestrategieën, die relatief gemakkelijk geïmplementeerd konden worden, vaker toegepast dan gespecialiseerde strategieën voor het vervullen van de specifieke onderwijsbehoeften van subgroepen of individuele leerlingen. In subgroepinstructies aan laagpresterende leerlingen werden soms aanbevolen strategieën voor rekenzwakke leerlingen toegepast, zoals het werken op een minder abstract handelingsniveau. Het was echter niet altijd duidelijk of deze aanpassingen aansloten bij de onderwijsbehoeften van de leerlingen, omdat er relatief weinig aandacht leek te zijn voor het diagnosticeren van onderwijsbehoeften op een meer kwalitatieve of geïndividualiseerde manier, bijvoorbeeld met behulp van diagnostische gesprekken. Het gebruik van meer kwalitatieve strategieën gericht op zowel het vaststellen van de specifieke onderwijsbehoeften van (individuele) leerlingen als het vervullen van deze specifieke onderwijsbehoeften door middel van aanpassingen in de instructie en verwerking is dan ook een tweede verbeterpunt.

Doel 3: Effecten van een nascholingstraject over differentiatie

Op basis van de specificatie van differentiatie is, in samenwerking met de experts in het consortium, een nascholingstraject ontwikkeld over differentiatie in het rekenonderwijs. In het nascholingstraject stonden de differentiatiecyclus en de bijbehorende differentiatiestrategieën centraal. Het nascholingstraject duurde één schooljaar en bestond uit tien teambijeenkomsten voor alle leerkrachten van de deelnemende scholen. Daarnaast zijn per school minimaal twee projectcoaches opgeleid om hun collega's, niet alleen tijdens maar ook na afloop van het nascholingstraject, te ondersteunen in het implementeren van differentiatie.

In hoofdstuk 3 zijn de effecten van het nascholingstraject op de geobserveerde implementatie van differentiatie door de leerkracht geëvalueerd. Hiertoe zijn in een deelsteekproef van 55 leerkrachten in Jaar 1 en 59 leerkrachten in Jaar 2 video-opnames van twee rekenlessen aan het begin en het einde van het schooljaar gescoord met het daarvoor ontwikkelde video-observatieinstrument (zie doel 2 en hoofdstuk 3). In Jaar 1 konden geen observeerbare effecten van het nascholingstraject op de implementatie van differentiatie worden vastgesteld: leerkrachten van de verschillende cohorten scoorden vergelijkbaar op zowel de voor- als nameting. Ook in Jaar 2 waren geen kortetermijneffecten van de interventie zichtbaar (leerkrachten van Cohort 2, die in dat jaar het nascholingstraject volgden, scoorden niet significant hoger aan het eind van het schooljaar dan aan het begin van het schooljaar). Er waren echter wel aanwijzingen voor een langetermijneffect van het nascholingstraject in Cohort 1. In Jaar 2 scoorden leerkrachten van Cohort 1, die in het voorgaande schooljaar het nascholingstraject hadden gevolgd, hoger op de implementatie van differentiatie dan leerkrachten uit de beide andere cohorten. Dit kon echter niet éénduidig worden geïnterpreteerd als een langetermijneffect van de interventie: de data lieten geen consistente stijging van de implementatie van differentiatie in Cohort 1 zien. In plaats daarvan leek er een algemene daling van de scores te zijn tussen Jaar 1 en Jaar 2, met name in Cohort 2 en Cohort 3. Diverse methodologische beperkingen, waaronder de relatief kleine deelsteekproef van geobserveerde leerkrachten en de uitdaging van het betrouwbaar meten van veranderingen in implementatie van differentiatie met behulp van een nieuw ontwikkeld video-observatieinstrument, zouden kunnen verklaren waarom het aantonen van *veranderingen* in implementatie van differentiatie lastig was. In de evaluatievragenlijsten, die zijn afgenomen na afloop van het nascholingstraject, rapporteerden leerkrachten dat ze het nascholingstraject wel als nuttig ervoeren en dat zij het geleerde toepasten in hun rekenlessen (zie hoofdstuk 4). In hoofdstuk 6 is dan ook geopperd dat de onderzoeksopzet van project GROW mogelijk sensitiever was voor het detecteren van veranderingen in rekenprestaties dan voor het detecteren van veranderingen in geobserveerd leerkrachtgedrag.

In hoofdstuk 4 zijn de effecten van het nascholingstraject op de rekenprestaties van de leerlingen geëvalueerd (*N* = 5658 leerlingen van groep 3 tot en met 8). In Jaar 1 vertoonden de leerlingen van Cohort 1, van wie de leerkrachten in dat jaar deelnamen aan het nascholingstraject, méér groei in rekenprestaties dan leerlingen van de controlecohorten (na correctie voor individuele verschillen in rekenniveau, non-verbale intelligentie en werkgeheugen). Dit effect was klein maar positief en dit gold voor zowel laag-, gemiddeld- als hoogpresterende leerlingen. Leerlingen van diverse prestatieniveaus hebben dus in vergelijkbare mate geprofiteerd van het nascholingstraject. In Jaar 2 was de groei in rekenprestaties echter ongeveer even groot in alle cohorten, dus leerlingen van Cohort 2 (van wie de leerkrachten in dat jaar het nascholingstraject volgden) vertoonden in de loop van het jaar niet méér vaardigheidsgroei dan leerlingen uit de andere cohorten. In hoofdstuk 4 zijn enkele praktische factoren benoemd waardoor leerkrachten van Cohort 2 mogelijk relatief veel aandacht moesten besteden aan andere zaken zoals het opbouwen van een nieuwe routine op een nieuwe school of het werken met een nieuwe rekenmethode. Hierdoor hadden leerkrachten van Cohort 2 mogelijk minder tijd en aandacht voor het nascholingstraject. Dit zou kunnen verklaren waarom de positieve effecten van het nascholingstraject die in Cohort 1 werden gevonden niet gerepliceerd konden worden in Cohort 2.

Doel 4: Relaties tussen rekenprestaties en rekenmotivatie

In hoofdstuk 5 zijn de relaties tussen rekenprestaties en verschillende aspecten van motivatie voor rekenen onderzocht met speciale aandacht voor mogelijke verschillen tussen laag-, gemiddeld- en hoogpresterende leerlingen. Dit is relevant, omdat de leerervaringen van leerlingen met verschillende prestatieniveaus behoorlijk kunnen verschillen (leerlingen met een laag prestatieniveau kunnen bijvoorbeeld relatief minder vaak succeservaringen opdoen bij het vak rekenen).

De resultaten van de totale steekproef (*N* = 4306 leerlingen van groep 4 tot en met 8) laten zien dat eerdere rekenprestaties een significante voorspeller zijn van motivatie, met positieve effecten op competentiebeleving en waardering voor het vak rekenen en een negatief effect op rekenangst. Bovendien hangen de verschillende aspecten van motivatie met elkaar samen. Kinderen die in het verleden goed hebben gepresteerd, voelen zichzelf dus meer competent op het gebied van rekenen en waarderen dit vak ook meer, terwijl zij juist minder rekenangst ervaren. Het effect van eerdere rekenprestaties op competentiebeleving reflecteert waarschijnlijk (gedeeltelijk) een correcte zelfinschatting van de eigen prestaties. Echter, er is ook een positief effect van competentiebeleving op latere rekenprestaties gevonden, zelfs na controle voor eerdere rekenprestaties. Bij vergelijkbare vroegere rekenprestaties hebben kinderen die hun eigen rekencompetentie hoger *inschatten*, dus een grotere kans om later beter te presteren. Hierbij is sprake van gedeeltelijke mediatie: eerdere rekenprestaties hadden een klein maar significant indirect effect op latere rekenprestaties via veranderingen in competentiebeleving. Voor de overige gemeten aspecten van motivatie – rekenwaardering en rekenangst – zijn geen significante effecten op latere rekenprestaties gevonden na controle voor eerdere rekenprestaties.

De bovenstaande patronen zijn grotendeels vergelijkbaar gebleken voor laag-, gemiddeld- en hoogpresterende leerlingen. Alleen bij de effecten van eerdere rekenprestaties op motivatie zijn enige verschillen tussen de groepen gevonden: bij hoogpresterende leerlingen waren de effecten van eerdere rekenprestaties op competentiebeleving en rekenwaardering sterker. De effecten van motivatie op latere rekenprestaties waren vergelijkbaar voor laag-, gemiddeld- en hoogpresterende leerlingen. Het relatieve belang van de verschillende motivatie-aspecten is in alle groepen ongeveer hetzelfde gebleken: competentiebeleving was voor zowel laag-, gemiddeld- als hoogpresterende leerlingen het enige motivatie-aspect met een positief effect op latere rekenprestaties, na controle voor eerdere rekenprestaties.

Algemene conclusies en implicaties

In hoofdstuk 6 zijn de resultaten van alle studies in dit proefschrift samengevat en bediscussieerd in relatie tot de vier gestelde doelen, met aandacht voor de beperkingen en implicaties van het onderzoek. Suggesties voor vervolgonderzoek zijn gedaan aan de hand van de volgende thema's: (1) het spanningsveld tussen enerzijds de wenselijkheid van differentiatiestrategieën vanuit het oogpunt van het vervullen van onderwijsbehoeften en anderzijds de praktische haalbaarheid van het toepassen van deze strategieën in de Nederlandse basisschoolcontext, (2) de uitdaging van het observeren van differentiatie, met name het meten van de mate waarin onderwijsaanpassingen daadwerkelijk afgestemd zijn op vastgestelde onderwijsbehoeften en (3) de effecten van differentiatie met behulp van niveaugroepen op de motivatie en in het bijzonder de competentiebeleving van leerlingen. Tot slot zijn de volgende algemene conclusies en implicaties opgesteld.

Differentiatie is een complex begrip. Het toepassen van differentiatie omvat méér dan enkel het werken met niveaugroepen: het vaststellen van onderwijsbehoeften, het stellen van gedifferentieerde doelen, het differentiëren van instructie, het differentiëren van verwerkingstaken, en het evalueren van de voortgang en het leerproces van de leerlingen. De differentiatiecyclus en de aanbevolen strategieën voor differentiatie bieden leerkrachten en lerarenopleiders concrete handvatten voor het toepassen van differentiatie in het rekenonderwijs. De meeste leerkrachten implementeerden al enige aspecten van differentiatie, maar bepaalde differentiatiestrategieën werden nog relatief weinig toegepast. De resultaten van dit proefschrift bieden daarmee specifieke aanknopingspunten voor het verbeteren van de implementatie van differentiatie. Ten eerste zouden verfijnde strategieën voor het diagnosticeren van (individuele) onderwijsbehoeften en specifieke aanpassingen van instructie en verwerking om deze vastgestelde onderwijsbehoeften te vervullen meer en beter gebruikt kunnen worden. Ten tweede zou meer aandacht kunnen worden besteed aan de behoeften van hoogpresterende leerlingen, met name op het gebied van (subgroep-)instructie. De resultaten met betrekking tot het nascholingstraject laten zien dat nascholing over differentiatie in het rekenonderwijs een positief effect kan hebben op de rekenprestaties van leerlingen met uiteenlopende rekenniveaus. Zulke effecten zijn echter niet gegarandeerd. In de dynamische context van het Nederlandse basisonderwijs is het een hele uitdaging om voldoende tijd en aandacht te vinden voor enerzijds nascholing over differentiatie en anderzijds de toepassing van het geleerde in de dagelijkse lespraktijk. De resultaten met betrekking tot de verbanden tussen rekenprestaties en motivatie lieten zien dat competentiebeleving bij leerlingen van alle prestatieniveaus het enige motivatie-aspect was dat niet alleen voorspeld werd door eerdere rekenprestaties, maar dat ook een significante voorspeller was voor latere rekenprestaties. Deze bevindingen impliceren dat het beïnvloeden van de competentiebeleving mogelijk een ingang is om ook de daadwerkelijke rekenprestaties, alsmede het plezier van de leerling in rekenen, te beïnvloeden. Bovendien verdienen de effecten van differentiatie met behulp van niveaugroepen op de competentiebeleving van leerlingen van verschillende prestatieniveaus nadere aandacht in vervolgonderzoek.

Dankwoord
Acknowledgements

Het schrijven van een proefschrift verloopt nooit precies zoals je had gepland, maar bij mij was de afwijking wel groter dan twee standaarddeviaties. Door een verkeersongeluk met langdurige gevolgen vergde het afmaken van dit proefschrift veel doorzettings- en aanpassingsvermogen, en dat gold zeker niet alleen voor mij. Ook mijn begeleiders en collega's, vrienden, familie: zo ongeveer iedereen die ik ken heeft zich in de afgelopen jaren moeten aanpassen aan een lastige situatie. Ik wil alle mensen die er voor me zijn geweest (en nog steeds zijn) dan ook hartelijk bedanken voor hun steun. Het is niet alleen aan jullie te danken dat ik dit proefschrift toch af kon maken maar ook dat ik het lachen niet heb verleerd.

Eva, Evelyn en Hans, jullie waren een geweldig betrokken begeleidingsteam in goede en in slechte tijden. Eva, je bent een ontzettend fijne persoon om mee samen te werken en als ik maar een kwart van jouw organisatievaardigheden heb afgekeken ben ik al aardig op weg. Ook de lezer van dit proefschrift mag jou dankbaar zijn, want jij behoedde me voor al te bloemrijk taalgebruik. Evelyn, met jouw scherpe geest en gulle lach heb je enorm bijgedragen aan dit project. Zwakke plekken in een betoog detecteerde jij meteen, wat de kwaliteit van dit proefschrift ten goede is gekomen. Ook was het altijd erg leuk om met jou te kletsen over gedeelde interesses. Hans, dankjewel voor je vertrouwen en betrokkenheid op het academische én het persoonlijke vlak. Je wist geldschieters ervan te overtuigen dat het heus de moeite waard was om te investeren in extra ondersteuning en verlenging van mijn contract, en jouw kwaliteiten als wandelende APA-manual zijn onbetaalbaar. Ik vond het een bijzonder teken van betrokkenheid dat je ook kwam luisteren als ik een concertje gaf met mijn strijkkwartet. Nu dit proefschrift dan toch eindelijk af is, mis ik onze projectbesprekingen. Ik mag jullie alle drie erg graag en ik hoop dan ook dat we contact blijven houden.

Aan de leden van de leescommissie – Eddie Denessen, Liesbeth Kester, Paul Leseman, Katrien Struyven en Jan van Tartwijk – veel dank voor het beoordelen van mijn manuscript en voor jullie bereidheid om hier met mij over van gedachten te wisselen tijdens de verdediging.

Mijn bijzondere dank gaat uit naar de consortiumleden en trainers: Bronja Versteeg van Rekenkracht (voorheen Giralis), Carla Compagnie-Rietberg, Lourens van der Leij en Martine van Schaik van het Marnix Onderwijscentrum, Jarise Kaskens en Anton Boonen van de Hogeschool Windesheim, Henk Logtenberg en Suzanne de Lange van CPS Onderwijsontwikkeling & Advies, Ina Cijvat, Tessa Egbertsen, Gert Gelderblom, Marcel Schmeier en Frans Winkels van Expertis, Marcel Absil, Ruud Janssen en Lenie van den Bulk van de CED-groep, Mieke van Groenestijn en Marianne Konings van de Hogeschool Utrecht en Karel Stokking en Gijsbert Erkens van de Universiteit Utrecht. Dankzij jullie expertise op het gebied van rekenonderwijs, differentiatie en nascholing van leerkrachten kreeg dit project een vliegende start. Jullie hebben een belangrijke en waardevolle rol gespeeld bij zowel de ontwikkeling van de training als de uitvoering ervan. Anneke Noteboom en Sylvia van Os van SLO, ook aan jullie veel dank voor het ruimhartige delen van trainingsmaterialen.

Tanja, Elien en Merel, als vaste onderzoeksmedewerkers hebben jullie een gigantische bijdrage geleverd aan het slagen van dit project. Soms waren de werkzaamheden stressvol, soms juist behoorlijk saai, maar jullie deden het allemaal met verve. Merel (nog een keer) en Rianne, bedankt voor jullie monnikenwerk bij het scoren van de video's. Elise, dankjewel voor jouw hulp bij de literatuurstudie over motivatie. Ook alle overige onderzoeksassistenten en studenten die in het kader van hun scriptie meewerkten aan de dataverzameling wil ik hier hartelijk bedanken voor hun inzet. Jullie hebben ontzettend geholpen bij het verzamelen van deze enorme berg data.

Uiteraard waren ook de deelnemende scholen onmisbaar. Ik wil bij dezen alle leerkrachten, projectcoaches, directeuren en niet te vergeten de leerlingen heel hartelijk bedanken voor hun medewerking bij de dataverzameling en voor hun inzet om de training tot een succes te maken.

Rens, ondanks jouw volle agenda wist je altijd wel ergens een gaatje te vinden om even met mij mee te denken over cryptische foutmeldingen in M*plus* of hardnekkige negatieve residuele varianties, dankjewel daarvoor. Milica, thank you for the pleasant collaboration on the motivation article.

Kamergenoten Sylke, Martine, Jaccoline, Ilona, Linda, Sietske en Rianne, dankjulliewel voor de gezelligheid en voor jullie luisterende oor op moeilijke momenten, maar ook voor de rustige werkomgeving die ik zo nodig had. Ook aan alle andere leden van de onderzoeksgroep veel dank voor de gezellige kletspraatjes op de gang en de warme belangstelling voor mijn wel en wee.

Anouk en Ineke, het geeft me veel vertrouwen dat jullie achter me zullen staan tijdens de verdediging. Anouk, het klikte altijd al goed tussen ons (en hoe kon het ook anders met zoveel gedeelde interesses) maar in de afgelopen jaren heb je echt mijn hart gestolen als ongelofelijk lieve en attente vriendin. Ineke, trouwe vriendin van altijd, wat heerlijk dat we ondanks diverse geografische scheidingen nog steeds zulk goed contact hebben. Janneke, we brengen niet meer zoveel tijd met elkaar door als toen we huisgenootjes én bestuursgenootjes waren, maar ik kan nog steeds alles met je delen en jij bent dan ook mijn officieuze derde paranimf. Aan jullie alle drie ontzettend veel dank voor de gezelligheid, de goede gesprekken en niet te vergeten de praktische (verhuis-)hulp in de afgelopen jaren.

Anne-Floor, Liselotte en Runa, de muzikale uurtjes met ton sur ton waarin niet alleen gespeeld maar ook veel gelachen werd fleurden me vaak helemaal op. Dankjulliewel dat jullie vanuit alle uithoeken van Nederland en België mijn huisje toch steeds weer weten te vinden om muziek te maken of (tegenwoordig steeds vaker en eigenlijk net zo leuk) gewoon bij te kletsen.

Ook aan mijn andere vrienden en familieleden, onder wie bestuursgenootjes van Cordon Bleu en andere Schuiters, Willemijn, oud-huisgenoten onder wie Simone en Anne, muziekvrienden, de Scheltensjes en iedereen die hier niet bij naam genoemd kan worden, veel dank voor de gezelligheid in de vorm van bestuursweekenden, avondjes bankhangen, legendarische Sinterklaas- en kerstvieringen, borrels, muziekavonden en al die andere leuke dingen. Cristina, thank you for being such a good friend. I have great memories of our trips and visits and hope we will spend many more cosy weekends together. Excellence friends, we don't get to see each other that often anymore but I am always delighted when we do.

Margreet, Gerald en Aldert, volgens goed gebruik bedank ik jullie hier als laatste terwijl (of omdat) ik heel veel aan jullie te danken heb. Aldert, zolang als ik me kan herinneren hebben we veel lol met elkaar gehad, of we nu samen op de ijsbaan stonden of op een podium. Dankjewel voor de hand- en spandiensten die je hebt verricht aan dit proefschrift, maar bovenal dankjewel dat je mijn lieve tamelijk groot uitgevallen kleine broertje bent. Margreet en Gerald, jullie hebben me altijd gestimuleerd om me te ontwikkelen in alle richtingen die ik maar wilde en dat heeft me gebracht waar ik nu ben. Jullie gaven me de ruimte om al jong op eigen benen te staan maar het is ontzettend fijn dat ik op jullie terug kan vallen als het leven even tegenzit. Dankjulliewel voor alles.


Mplus
User's Guide
Hush Puppies
REPLAY

About the author

Curriculum vitae

Emilie Prast was born on December 31st, 1988, in Leiden (the Netherlands). She obtained her secondary school degree (gymnasium) from the Rijnlands Lyceum in Oegstgeest. In 2005, she enrolled at Leiden University to study Education and Child Studies, combined with a minor in violin playing (Practicum Musicae) at the Royal Conservatory of The Hague. Besides, Emilie participated in two interdisciplinary honours classes and was actively involved in the student sailing association (president of the board in 2007–2008). After obtaining her bachelor's degree in 2009, she moved to Munich to pursue her interest in giftedness and talent development at Ludwig-Maximilans-University. During the two-year master's programme in educational psychology with a focus on gifted education, her academic interest broadened to education as a means to develop the potential of *all* children. She continued to explore this topic in her PhD project about differentiated instruction for students of diverse achievement levels in mathematics, which she commenced at Utrecht University in 2011 in collaboration with Eva Van de Weijer-Bergsma and under supervision of Hans van Luit and Evelyn Kroesbergen. Emilie played an important role in the design of the teacher professional development programme which was evaluated in this large-scale project. Moreover, she developed new instruments to measure the implementation of differentiation as well as students' motivation for mathematics. She presented her work at various conferences, including a keynote speech at the EARLI SIG 11 conference in 2016. Emilie was also involved in educational activities including master thesis supervision and lecturing. Since January 2018, Emilie works at Leiden University as an assistant professor in educational sciences.

Research output

Publications in international scientific journals:

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2015). Readiness-based differentiation in primary school mathematics: Expert recommendations and teacher self-assessment. *Frontline Learning Research*, *3*, 90-116. doi:10.14786/flr.v3i2.163.

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2018). Differentiated instruction in primary mathematics: Effects of teacher professional development on student achievement. *Learning and Instruction*, *54*, 22-34. doi:10.1016/j.learninstruc.2018.01.009

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Prast, E. J., & Van Luit, J. E. H. (2015). Validity and reliability of an online working memory task for self-reliant administration in school-aged children. *Behavior Research Methods*, *47*, 708-719. doi:10.3758/s13428-014-0469-8

Manuscripts submitted for publication in international scientific journals:

Prast, E. J., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (submitted). Teaching students with diverse achievement levels: Observed implementation of differentiation in primary mathematics education.

Prast, E. J., Van de Weijer-Bergsma, E., Miočević, M., Kroesbergen, E. H., & Van Luit, J. E. H. (submitted). Relations between mathematics achievement and motivation in students of diverse achievement levels.

Publications for educational professionals (in Dutch):

Prast, E. J., Van Hooijdonk, M., Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2013). Wat vinden leerlingen van hun rekenlessen? Rekenen met plezier. *Jeugd in School en Wereld (JSW)*, *98*(2), 6-9.

Van de Weijer-Bergsma, E. & Prast, E. J. (2013). Gedifferentieerd rekenonderwijs volgens experts: De resultaten uit een Delphi onderzoek. *Orthopedagogiek: Onderzoek & Praktijk*, *52*, 336-349.

Van de Weijer-Bergsma, E., Prast, E. J. Kroesbergen, E. H., & Van Luit, J. E. H. (2012). Afstemmen op onderwijsbehoeften: Gedifferentieerd rekenonderwijs. *Volgens Bartjens*, *31*(4), 31-33.

Van de Weijer-Bergsma, E., Van Luit, J. E. H., Prast, E. J., Kroesbergen, E. H., Kaskens, J. M. M., Compagnie-Rietberg, C. W., Cijvat, I., & Logtenberg, H. (2016). *Differentiëren in het rekenonderwijs: Hoe doe je dat in de praktijk?* Doetinchem: Graviant scientific & educational books.

Presentations at national and international conferences:

Prast, E.J. (2016, June). *Differentiated instruction in primary mathematics: Effects of teacher professional development on teachers' instructional behaviour and student achievement*. Invited keynote speech presented at the biennial EARLI SIG 11 conference, Zurich, Switzerland.

Prast, E.J. (2015, November). *Effects of readiness-based differentiation on student achievement in primary mathematics education*. Paper presented at the ISED-VNOP-CAS Research Days, Leiden, The Netherlands.

Prast, E.J., Van de Weijer-Bergsma, E., Kroesbergen, E.H., & Van Luit, J.E.H. (2015, May). *Effecten van een nascholingstraject over differentiatie in het rekenonderwijs op de rekenprestaties van basisschoolleerlingen*. Paper presented at the Onderwijs Research Dagen, Rotterdam, The Netherlands.

Prast, E.J., Kroesbergen, E.H., Van de Weijer-Bergsma, E., Van Luit, J.E.H., & Van Hooijdonk, M. (2015, August). *Enhancing quality of math instruction with teacher professional development about differentiation*. Paper presented at the biennial EARLI conference, Limassol, Cyprus.

Prast, E.J., Van de Weijer-Bergsma, E., Kroesbergen, E.H., & Van Luit, J.E.H. (2013, August). *Teacher self-assessment of differentiation strategies recommended by experts*. Paper presented at the biennial EARLI conference, Munich, Germany.

Prast, E.J., Van de Weijer-Bergsma, E., Kroesbergen, E.H., & Van Luit, J.E.H. (2013, August). *What constitutes good differentiation in primary mathematics education? Results of an expert consensus procedure*. Paper presented at the biennial JURE conference, Munich, Germany.

Measurement instruments developed for project GROW:

Differentiation Self-Assessment Questionnaire (DSAQ)*
Questionnaire for teachers about their perceived implementation of a range of strategies for differentiation in primary mathematics education. Described in Chapter 2 of this dissertation.

Differentiation in Mathematical Instruction (DMI)*
Observation instrument for observing teachers' implementation of a range of strategies for differentiation in videotaped primary mathematics lessons. Described in Chapter 3 of this dissertation.

Mathematics Motivation Questionnaire for Children (MMQC)*
Questionnaire for primary school children about various aspects of motivation for mathematics. Described in Chapter 5 of this dissertation.

The Lion game**
Visual-spatial complex span task for self-reliant administration in primary school children. Described in: Van de Weijer-Bergsma, E., Kroesbergen, E. H., Prast, E. J., & Van Luit, J. E. H. (2015). Validity and reliability of an online working memory task for self-reliant administration in school-aged children. *Behavior Research Methods*, *47*, 708-719. doi: 10.3758/s13428-014-0469-8

The Monkey game**
Backward word span task for self-reliant administration in primary school children.
Described in: Van de Weijer-Bergsma, E., Kroesbergen, E. H., Jolani, S., & Van Luit, J. E. H. (2016). The monkey game: A computerized verbal working memory task for self-reliant administration in primary school children. *Behavior Research Methods*, *48*, 756-771. doi:10.3758/s13428-015-0607-y

* Emilie Prast was the primary developer of this measurement instrument
** Eva Van de Weijer-Bergsma was the primary developer of this measurement instrument