The Effects of Group differentiation by students' learning strategies

Carla Haelermans¹

Received: 16 July 2021 / Accepted: 13 December 2021 / Published online: 11 January 2022
© The Author(s) 2022

Abstract

This study analyses the effects of group differentiation by students' learning strategies of around 1200 students in 46 classes from eight secondary schools in the Netherlands. In an experimental setup with randomization at the class level, division of students over three groups per class (an instruction-independent group, an average group, and an instruction-dependent group) is based on learning strategies, measures using the Motivated Strategies for Learning Questionnaire (MSLQ). Each group is offered instruction fitting their own learning strategy. The results show that student performance is higher in classes where the differentiation was applied, and that these students score higher at some scales of the posttest of the questionnaire on motivation, metacognition and self-regulation. However, there are differences between classrooms from different teachers. Additional teacher questionnaires confirm the discrepancy in teacher attitudes towards the intervention.

Keywords Differentiation · Secondary school · Motivated strategies for learning · Randomized experiment

Introduction

Students enter the classroom with different abilities, personalities, learning strategies, attitudes and motivation (Vernooij, 2009). Learners differ in several aspects of culture, intelligence, language, economic background, gender, motivation and interests (Tomlinson et al., 2003). Taking into account these differences is expected to improve learning as humans are motivated to learn what matches their individual interests and deeply felt needs (George, 2005). In the Dutch educational system, as well as in many other countries, these differences are partially tackled by external differentiation on the meso level, meaning that students are tracked based on cognitive level, social-emotional skills, and interests. However, there is

✉ Carla Haelermans
Carla.haelermans@maastrichtuniversity.nl

¹ Research Centre for Education and the Labour Market, School of Business and Economics, Maastricht University, PO box 616, 6200MD Maastricht, the Netherlands

still a lot of heterogeneity within tracks. Furthermore, recent international developments such as inclusive education and dismissal of gifted programs (Sapon-Shevin, 2000/2001), obliges teachers to introduce differentiated instruction in the regular heterogeneous classroom as well, instead of providing organizational arrangements (Jackson & Davis, 2000). Differentiation is broadly defined by, for example, Tomlinson (2000) as “tailoring instruction to meet individual needs”, but in the literature many definitions are found, both broad and more narrow (see Roiha & Polso, 2021 for an overview). Providing differentiated classroom instruction, i.e. the adaptation of classroom strategies to students’ different learning interests and needs so that all students experience challenge, success, and satisfaction, is one type of differentiation which enables to respond to this diversity in students’ interests and needs, which is absolutely essential in education (George, 2005). George (2005) states that heterogeneous classrooms with differentiated instruction prepare students better for real-life situations, now and in the future, achieving new roles and relationships and lead to significant learning that is personally meaningful, satisfying, transferable and long-lasting (Tomlinson, 2000).

However, teachers often do not use differentiated instruction in the classroom, for various reasons. Teachers for example often find it difficult to provide all students with those learning activities that fit them best, or have a lack of teacher self-efficacy (Dixon et al., 2014). Melesse (2015), for example, conducted surveys, semi-structured interviews and focus groups with 232 primary school teachers in Ethiopia and concluded that many factors play a role in experiencing challenges of differentiation, such as knowledge and experience, class size, time availability and availability of resources. In a qualitative study among 322 Norwegian student teachers, Brevik et al. (2018) concluded that even though student teachers are aware of the importance of differentiation, they lack the confidence to use it. Brevik et al. (2018) emphasize the importance of teacher education to teach future teachers to effectively differentiate.

Within-classroom differentiation is in this study defined as “*an approach to teaching in which teachers proactively modify curricula, teaching methods, resources, learning activities, and student products to address the diverse learning needs of individual students and small groups to maximize the learning opportunity for each student in the classroom*” (Tomlinson et al. (2003, pp 121) and can take place according to the student’s cognitive performance/readiness, a student’s interests or a student’s learning profile/learning strategy (Tomlinson & Moon, 2013). The most commonly used way to differentiate is based on cognitive performance. Dutch teachers who *do* provide some differentiated instruction mainly focus on differentiation based on cognitive level/performance level, most likely because this is the most known type of differentiation and the easiest to implement and group students on (van Casteren et al., 2017). Differentiated instruction based on cognitive level has mainly a positive effect on cognitive growth (Reis et al., 2011). For differentiated instruction based on students’ interests, positive correlations are found with creativity and motivation. Lastly, differentiation based on learning strategies (Dunn & Dunn, 1993) is found to correlate with cognitive growth (Lovelace, 2010), creativity (Kyprianidou et al., 2012), collaboration (Kyprianidou et al., 2012), and attitudes towards learning (Lovelace, 2010). This third way of differentiating is the focus of our study.

Differentiated instruction based on learning strategies seeks to maximize each student’s growth by acknowledging that students do not only differ in cognitive level or interests but also in learning profile, learning strategies and metacognitive skills (Ravitch, 2007 in

Lauria, 2010). As mentioned above, differentiation based on learning strategies correlates with cognitive growth, creativity, collaboration, and attitude towards learning. It is important to make clear that in this study we are not talking about learning styles (the way in which an individual works with, processes and internalizes information (Lauria, 2010)) but about learning strategies (metacognitive skills such as the strategy that the student applies to organize and plan and regulate the study effort, the ability of the student to think critically and the level of self-regulation of the student (Pintrich & de Groot, 1990)). Although some studies have found positive relations between learning style preference on outcomes (e.g. Lauria, 2010 and Dunn & Honigsfeld, 2013), Kirschner (2017) points out that there is no objective evidence that a student has a certain optimal learning style; that there is a valid and reliable way to measure/determine this style; or an optimal instructional method to align with this learning style. Learning strategies on the other hand can help students increase their self-regulation abilities and overall metacognition to improve learning. In the study at hand, the concept of learning strategies is defined within the framework of self-regulated learning and metacognitive strategies of Pintrich and de Groot (1990).

Metacognitive strategies and self-regulation

Metacognition refers to students' ability to monitor and control cognitive processes. It is typified as '*a person's cognition about cognition*', or as Swanson (1990, p. 306) stated "*the knowledge and control one has over one's thinking and learning activities.*" The term metacognition has many definitions depending on how it is used (e.g., Pintrich, 2000; Winne & Hadwin, 2008; Zimmerman, 2000), but an important commonality is that metacognition always refers to monitoring and controlling cognitive processes. Here, metacognition is defined as one's ability to regulate cognitive processes and is strongly related to learning outcomes (Coutinho, 2006; Flavell, 1979). It refers to higher-order thinking, and includes skills that enable learners to think about, understand, and monitor their learning (Schraw & Dennison, 1994). Students with more metacognitive skills are better equipped to take advantage in learning environments and metacognition therefore plays an important role in student learning (Dunlosky & Thiede, 1998; Dunning et al., 2003). Students' learning allows them to monitor and regulate cognitive activities in a way that improves learning outcomes (Nelson & Narens, 1990). Metacognitive skills can be improved in classrooms by creating a learning environment in which students demonstrate, explain, discuss, and control their own thought processes. To do this, teachers need to engage students in metacognitive activities, such as tasks that stimulate them to think about how to address problems while completing these tasks. Metacognitive strategies help planning, monitoring and modifying a student's cognition. Four metacognitive strategies are defined by Schraw (1998): orienting, planning, monitoring and evaluation. Orienting makes sure previously acquired knowledge is activated and the goal of the assignment was determined. Self-efficacy plays an important role in this phase. Planning means selecting the appropriate strategy and calculating needed time and effort. Monitoring requires some awareness of the process of understanding and executing the task. In the process of evaluating, the student assesses the product and the process.

A central aspect in metacognition is self-regulated learning, which is a process whereby students set goals for their learning and where they monitor, regulate and control actions, cognition and motivation needed to achieve these goals (Zimmerman & Labuhn, 2012; Schraw et al., 2006). As part of self-regulated learning, *cognitive strategies* like rehearsal,

elaboration and organizing will help students to reach their cognitive goals. Cognitive strategies are used to stimulate different kinds of information processing. With simple tasks, rehearsing can be a good strategy to process information, but it will not support the process of connecting different pieces of information or assimilating new information with previously acquired knowledge. Organizing is restructuring information to transform the new information in meaningful units and transfer it to long-term memory. Organizing stimulates the process of selecting information and making connections between pieces of information. Organization demands time and effort, but it is believed to increase students' involvement with the task at hand. Elaboration is used to anchor the information in order to connect the new information with previously learned information. Elaboration is the process of extensively processing information by describing, summarizing, note-taking and finding analogies (Pintrich et al., 1991).

Within-classroom differentiation based on self-regulation and metacognitive skills

In addition to what we have described above, the theoretical background of within-classroom differentiation is nicely described in Denessen and Douglas (2015). The authors discuss that within-classroom differentiation is needed to address diverse learning needs in order to provide an optimal learning context for all students (Roiha 2014; Sarrazin et al., 2006), whereas whole-classroom teaching is a less-attractive pedagogical practice (Tomlinson et al., 2003). As discussed above, within-classroom differentiation can be applied based on different student characteristics (cognitive performance/readiness, interests and learning profile/learning strategies). Furthermore, teachers can differentiate through content, process, product or affect/environment (Tomlinson & Moon, 2013), for example using different tasks at different levels or a different pace of instruction (Tomlinson et al., 2003 in Denessen & Douglas, 2015). Denessen and Douglas discuss that these types of differentiated instruction are grounded in constructivist theories, such as Vygotsky's theory on learning and teaching. According to Subban (2006), Vygotsky's theory on the zone of proximal development is the basis for differentiated instruction. Differentiation challenges students to bridge the gap between their current performance and their potential performance.

As Denessen and Douglas mention in their chapter, the teacher has different options to differentiate, such as given additional instruction to some students, letting students work independently or give more challenging assignments. However, these types of differentiation are only likely to work effectively if the students have the metacognitive and self-regulation skills to, for example, work independently or deal with more challenging assignments. Based on this, differentiation by learning strategies is a good alternative to differentiation solely based on cognitive performance, as there is no automatic relation between cognitive performance and learning strategies (Hames & Baker, 2015), implying that mere cognitive differentiation is not sufficient or may not be applicable at all. Therefore, depending on the type of instructional differentiation activities the teacher wants to use, creating groups based on learning strategies may be preferable.

However, these differentiation approaches are only effective, meaningful, and adaptive if they match individual students' learning profile/learning strategies (Denessen & Douglas, 2015). Therefore, the decision of the teacher to offer which students what type of differentiate instruction should be dependent on good knowledge by the teacher on how the students differ on these learning strategies. As teachers are often biased in their assessment of stu-

dent's cognitive level (especially in the case of students from a disadvantaged background or students with an ethnic background (see e.g. Ready & Wright, 2011)), and often wrongfully use student's cognitive level as an indication for learning strategies, it is important to use external instrument(s) to measure students' cognitive level/learning profile/learning strategies.

Note that there is no theoretical or empirical ground to argue that students have only one learning strategies for all subjects and learning activities (Lovelace, 2005), implying that it is important to explicitly take into account that measuring metacognitive and self-regulated strategies in order to determine learning strategies is very subject and situation specific.

Research questions

As mentioned above, there are only a few studies that look into the effect of differentiated instruction based on learning strategies (see Lovelace, 2010 and Kyprianidou et al., 2012) and there is not much knowledge yet on the effect of differentiation by learning strategies on student performance and on motivation and metacognitive and self-regulation skills. Therefore, this study focused on the effect of differentiated instruction based on learning strategies on learning performance and on metacognitive and self-regulation strategies and motivation. The goal of this research is to find an answer on the following four questions:

1. What is the effect of differentiated instruction by learning strategies on student motivation, metacognition and self-regulation?
2. What is the effect of differentiated instruction by learning strategies on student performance?
3. (How) do the found effects differ between the three groups based on learning strategy?
4. Do the teachers perceive the differentiated instruction process as feasible and valuable?

Methods

The intervention

In this research, the effect of differentiated instruction based on students' learning strategies is studied. The intervention took place in the school year 2015/2016. In the experimental classes, students were divided in three groups, based on five scales of the MLSQ (which is described in detail later in this paper) from the study skills and learning strategies part: organization, elaboration, metacognitive self-regulation, time and study environment, and effort regulation. These five scales were selected as they are most closely related to the intervention at hand. The other scales from the study skills and learning strategies part of the Motivated Strategies for Learning Questionnaire, hereafter MSLQ, did not relate to the intervention, content wise. Based on the mean scores on these scales, the students were divided in three groups: an instruction-independent group, an average group, and an instruction-dependent group. As a large part of the students (on average 50%) scored between 4 and 4.3 (on a 7-point Likert scale), we used these as the upper and lower boundary to define the groups. Students were categorized as instruction-independent when their mean score on all five scales was higher than 4.3, and students were part of the instruction

dependent group if their average score on at least four out of five of these scales was lower than 4. The remaining students were categorized in the average group. This way of creating groups leads to an unequal division of students over groups, which in most cases leads to a large average group, a smaller group that needs teacher instruction and guidance, and a very small group that worked mostly independent. Furthermore, this led to homogeneity between the upper and lower group, but not necessarily in the middle group. However, the teachers also had some discretion to switch children between groups, to also take into account readiness and knowledge of the student to work on the specific material in the specific group that the student was put in (mostly students with low cognitive performance in the instruction independent group). Although teachers were not specifically asked to inform us about every individual switch they did, many of them informed us about this, and these switches led to a better homogeneity within groups, as well as heterogeneity between groups. Note that, although it is likely that students from different groups talk to each other about the differences in instruction that they receive, it is not very likely that this will influence the treatment they receive, as the treatment is really the teacher instruction difference and the independence of students, and in that sense a difference in attention from the teacher for each group. The lower and the middle group get the same content, with the difference that the middle group works on assignment on their own whereas the lower group does that under guidance of the teacher. Only the upper group received additional assignments to work on independently. We believe that it is unlikely that students talking to each other will influence the treatment, but in the unlikely event that it would, that would mean that the differences between the treatments are smaller, implying that the results that we find are a lower bound of the 'real' results if there were no contamination between groups.

The lessons in which the intervention was carried out were developed according to the IDDI-model that stands for Interactive Differentiated Directed Instruction. This model is a variant of the direct instruction model that is found to be one of the most effective instruction methods (Camilli et al., 2010). Camilli et al.'s direct instruction model is characterized by the following teaching behaviors: Presentation and explanation of the to-be-learned materials; increasing interest of the students; presenting clear goals and guidelines; conscientious, structured practice and coaching; direct, correcting feedback; and monitoring understanding of the core theoretical aspects.

The direct instruction model assumes a very structured lesson plan, in which several phases of orientation, learning and reflection are alternated. In the differentiated instruction model, this model is extended with cooperative methods and convergent differentiation in groups, especially in the phases of instruction and practice. The lesson starts with a common introduction on the subject. All students receive a short instruction of the learning content after which the instruction-independent group starts working with the materials. The teacher then presents the rest of the group an interactive group instruction with an example on how to practice with the materials. The average group started to work with the materials while the instruction-dependent students received a prolonged instruction and guided practice. The teachers were trained in this model in two different meetings and additional support was available during the whole intervention period. Teachers were expected to take responsibility for finding or developing necessary materials but were given the opportunity to submit these materials to an expert in the differentiated instruction model for advice, as well as the opportunity to develop these materials together. The teachers were asked to perform this intervention in the classes which were divided in three groups by the researchers, to do this

minimally once a week and to report any deviations in procedure to the researchers so that they could control for these aspects in the analyses.

In relation to the five MLSQ scales that were used to divide students into the three groups, in this differentiated instruction model, the upper group (instruction-independent) did not only get fewer instructions, but were expected to organize their own work. Furthermore, they were expected to further elaborate the content of the topic they were studying by themselves (as the assumption was that they would need less time than the other two groups to go through the basic content), based on the learning materials that were made available by the teacher. And they were expected to regulate themselves and plan their time and study effect accordingly. The lower group, the instruction-dependent group, received a lot of guidance by the teacher, and the teacher planned and organized the learning content. This group worked only on the basic content of a topic, without further elaborating on related aspects. The average group was a mix of students, and the guidance by the teacher was accordingly. Depending on the strength and weaknesses of a student the teacher provided more or less guidance in organization, planning, elaboration or regulation.

Procedure

First, a pretest of the MSLQ (MSLQ, Pintrich et al. 1991) was given to divide the students in the experimental group over three groups per class. Furthermore, the pretest was used to control for potential group differences that might occur despite the randomization procedure. The average grade of the student for the subject in which the intervention took place (or for which the student was a control student) in the first period of the school year under study is the pretest measure of this study. Then, the intervention was implemented in one teaching unit per week during eight months of the school year. The grade in the period immediately after the intervention took place is the post-test measure, and one of the outcomes. The second outcome is the MSLQ, which allows for measuring both study approach (among which self-regulation skills) as well as study motivation at the same time. Finally, teachers were asked to fill in a questionnaire on both their beliefs about and experiences with the intervention to answer the question whether the intervention was feasible for teachers. The MSLQ and the questionnaire will be explained in more detail later.

Instruments and measures

Grades

In Dutch secondary education, grades range between 1 and 10, although grades are usually between 4 and 9 (grades below 4 and the highest grade 10 are rarely given). A 5.5 (55% of the total score) is sufficient to pass the exam or the course and the majority of grades is between 5 and 7. Although standardized tests and exams are used for example in graduation year for the national exams, in general tests are developed and graded at the school level, most often by the teacher him/herself. However, the school board to which all of the participating schools belong uses a slightly different approach. Here tests, especially the larger and more important tests that are conducted in test weeks, are usually developed with all teachers from the section of a specific subject for a given grade level and are graded by this same group of teachers. And in the case of our study, all participating teachers from the different

schools that teach the same subject have jointly developed the materials for the differentiated instruction for the treatment groups. This implies that the grades in our study are quite reliable, comparable and can be considered valid measures of students' actual performance.

MSLQ

In order to get information on the cognitive and meta-cognitive strategies the students use, as well as on their study motivation before and after the experiment, we used the Motivated Strategies for Learning Questionnaire (MSLQ). The MSLQ is a validated questionnaire that can be used in a flexible way, that has been used many times before and that allows one to measure motivation and learning strategies of students (Artino, 2005). The questionnaire was developed and first validated by Pintrich et al. in 1991. The MSLQ consists of 81 questions divided over two parts: 31 questions about motivation and 50 questions about study skills and learning strategies. In this study, an adaptation of the Dutch version of the MSLQ was used, which was translated by the Open University (van der Boom, 2009). This Dutch version was originally developed for university and college students; therefore, it was adapted to be used in this study of secondary school students, and we pilot tested the MSLQ in the schools that participated in this study in the year prior to the intervention (twice), using students that would for sure not participate in the study. The questionnaire was adapted after each test round, and mostly changed towards a better understanding of younger students and students of the lower, prevocational, track. We did not remove any statements, but only changed the wording of certain statements when students had difficulties understanding the statement. Students answer each question on a 7-point Likert scale, ranging from 'completely disagree' to 'completely agree'. The MSLQ can be divided into six sub scales for the motivation questions and nine sub scales for the study skills and learning strategies questions. Table 1 shows the two parts of the MSLQ, the 15 corresponding sub scales and the Cronbach's alpha per scale. The instrument MSLQ is used twice during the study at hand, once as a pretest at the beginning of the school year (September 2015), and once as a posttest after the experiment (April/May 2016). Additionally, the difference on each sub scale between the pretest and posttest is also used as an outcome measure, to see whether students develop certain skills and behavior differently when they were part of the intervention group.

The reliability of the questionnaire has been tested by the developers of the MSLQ (Pintrich et al., 1993), and by many others, who demonstrate that the MSLQ can be used across a variety of different samples (from primary to university education) in many different kinds of disciplines, with good confidence for obtaining generally reliable scores (e.g. Midgley et al., 1998; Erturan Ilker et al., 2014; Feiz et al., 2013 and Taylor, 2012). For our sample, reliability is checked using Cronbach's alpha and compared with the alphas from the original questionnaire by Pintrich et al. (1993). Table 1 shows this analysis. In most of the 15 cases our reliability measure is a little bit lower than the one from Pintrich et al. which could be due to the fact that we use a Dutch modification of the original version and that we apply the questionnaire to secondary instead of college students. Previous studies applying the MSLQ to high-school students both in the Netherlands and elsewhere have also shown a somewhat lower reliability on most of the scales (see e.g. Chow & Chapman, 2017 and Blom & Severiens, 2008). Note that the reliability of the scales hardly differs when we analyze them separately for subgroups of students by track or age group.

Table 1 Reliability analysis MSLQ

Scale	Subscale	Cronbachs alpha (Pintrich et al. 1991)	Cronbachs alpha (current study, n=1228)
Motivation	Intrinsic goal orientation	0.74	0.53
	Extrinsic goal orientation	0.62	0.55
	Task value	0.90	0.80
	Control beliefs	0.68	0.54
	Self-efficacy for learning and performance	0.93	0.90
	Test anxiety	0.80	0.80
Learning strategies	Rehearsal	0.69	0.75
	Elaboration	0.76	0.76
	Organization	0.64	0.76
	Critical thinking	0.80	0.69
	Metacognitive self-regulation	0.79	0.75
	Time and study environment	0.76	0.69
	Effort regulation	0.69	0.69
	Peer learning	0.76	0.59
	Help seeking	0.52	0.59

Note that we underlined the learning strategies on which the division of treatment students into groups is based

However, most scales have a reliability of 0.7 or higher, which is generally accepted as being reliable (Field, 2013). Only five sub scales have low reliability. These are intrinsic and extrinsic goal orientation, control beliefs, peer learning and help seeking. We have removed the unreliable scales from our analysis and from the remainder of this paper.

Teacher questionnaire

A teacher questionnaire was administered post-test only. The goal of this questionnaire was to check whether the teachers implemented the intervention as supposed and to gather some information on their experiences and satisfaction with the intervention. The teacher questionnaire consisted of 40 questions which had to be answered on a 5-point Likert scale. For each question, the teachers were given the opportunity to add a comment to explain their answers on the statements. The questionnaire was based on the expectancy theory of Shah and Higgins (1997). Abrami et al. (2010) used this theory to construct a model studying the factors that influence a teacher's decision to implement an educational innovation and persist in its use. According to them, a teacher's decision to implement an innovation is related to (1) how successful he/she expects to be in the implementation of this innovation, (2) how he/she values the innovation as worthwhile the effort, and (3) how high he/she perceives the costs. Based on this model, Abrami et al. (2010) developed a questionnaire to study implementation of cooperative learning, on which our questionnaire is based. Some items were adapted to fit our context and other items were deleted when not relevant. *Expectancy items* were formulated to study teachers' perceptions of the contingency between their use of the intervention and the desired outcomes and factors affecting these perceptions. In the expectancy scale, 16 items were formulated on teacher self-efficacy and skill (e.g., "I

do understand the principle of differentiated instruction well enough to implement it in my classroom”), student characteristics (e.g., “My students do not have the skills to work according to this model”), classroom environment (e.g., “The physical space I have, enables me to work with differentiated instruction”) and collegial and managerial support (e.g., “To implement this model, I have the support of my co-workers”). The *value scale* consisted of seven items, measuring the degree to which teacher perceived the innovation as worthwhile for the learning of the students. Example items are: “differentiated instruction is a valuable approach to provide for tailor-made solutions” and “differentiated instruction stimulates positive feelings about learning among the students”.

Finally, teachers were asked for their perceptions of the *physical and psychological demands* of implementation operating as disincentives for implementing the innovation in six items. These items represent class and preparation time, effort and specialized materials. An example item is: “Differentiated instruction demands too much preparation time for me as a teacher”. This questionnaire was supplemented with nine questions, serving as a manipulation check. We asked the teachers to point out when they started and ended the intervention, how many times they taught the experimental class, whether they adapted the categorization of the students in the three groups and whether they used the offered support of the expert. Additionally, they were given the opportunity to reflect on their experiences during the process by asking them whether they thought the intervention was a success or failure, which factors contributed to this success/failure and whether they were interested to continue this method.

Note that the teacher questionnaire was not tested for validity and reliability, so the results are merely indicative of how the teachers felt about the intervention. The specific items of the questionnaire can be found in Appendix 1.

Research design

An experimental design was used to study the effect of differentiated instruction based on students’ learning strategies on motivation and learning performance. Specifically, a comparison was made between an experimental group who received differentiated instruction, at least once a week for a period of eight months and a control group who received general instruction for the same period. Randomization took place at the class level, such that each teacher has at least one experimental and one control class. The randomization took place before the school year started and was executed by the researchers.

Sample and descriptive statistics

Descriptive statistics final sample

The final sample of this study consists of 1228 students in 46 classes from eight schools located in the midst of the Netherlands, all belonging to the same school board. The number of teachers and classes per schools varied between one teacher per school to nine teachers per school, and from two classes per school to 18 classes per school, with student numbers per school ranging between about 50 and 500. Table 2 shows the characteristics of the participants. The ability score in this table is based on the so-called CITO-test. This is an ability test written at the end of primary school that is, among others, used for track placement in

Table 2 Background characteristics students

Variable	Observations	Mean	Standard deviation	Minimum	Maximum
Age	1227	15.39	1.53	12	20
Gender (female = 1)	1228	0.54	0.49	0	1
Ability	960	541.33	6.17	516	550
Dutch	1228	0.81	0.4	0	1
Pretest grades	1161	6.78	1.2	2.7	9.7
Posttest grades	1147	6.55	1.12	1	9.6

secondary school. The CITO- test is presented in IRT-scores and ranges from 501 to 550. Although it is compulsory for primary schools to administer an end test, there are multiple test providers to choose from. However, at the time this research study took place, the large majority chose for the CITO-test, and secondary schools often did not administer test scores from other test suppliers. Therefore, we do not have the ability score available for all students in our sample, although we still have this information for about 80% of students. The variables for AD(H) and Dutch speaking are dummy variables, where the mean indicates the share of students that have AD(H)D/speak Dutch at home. Table 2 shows that participating students are on average 15 years old and have an average ability score of 541. A little more

Table 3 Background characteristics students

Variable		Amount	Percentage
Track advice	Prevocational education	292	24.11%
	Higher general education	544	44.92%
	Pre-university education	375	30.97%
	Total	1211	
Grade level	7	212	17.26%
	8	289	23.53%
	9	117	9.53%
	10	460	37.46%
	11	150	12.21%
	Total	1228	
Level	Prevocational education	153	12.46%
	Higher general education	536	43.65%
	Pre-university education	539	43.89%
	Total	1228	
Subject	Mathematics	296	24.10%
	Dutch	54	4.40%
	English	243	19.79%
	French	51	4.15%
	Physics	145	11.81%
	Economics	48	3.91%
	Management & Organisation	32	2.61%
	History	89	7.25%
	Biology	222	18.08%
	German	48	3.91%
	Total	1228	

than half (54%) are girls, and 81% are Dutch speaking. They have an average grade (measured between 1 and 10) of 6.78 in the pretest and 6.55 in the posttest.

Based on Table 3, it seems that mostly students in higher tracks participated in the study. Dutch secondary education consists of three tracks, the lowest being prevocational education (duration four years, age 12–16), the middle track is higher general education (five years, age 12–17) and the highest track prepares for university education (six years, age 12–18). After primary education, students are placed in a track based on the results of a standardized ability test at the end of primary school in grade 6 (in most cases the CITO-test, as explained above) and a placement advice (based on social-emotional skills and interests of the student, as well as the cognitive level) given by the 6th grade teacher in primary school.

About 75% of the students in our sample received a primary school track placement advice for the middle or the higher track, implying that they are in the upper half of the ability distribution. This is confirmed by the actual track of the students, both the middle and the higher track are overrepresented compared with the lower track. Table 3 furthermore shows that most students in this study were in grade level 10 at the time of the research, followed by grade 8 (American grading system used). Lastly, it is shown that the intervention mostly took place in mathematics classes, followed by English and Biology. Note that this was not a choice of the students, but determined by the teacher, as it was the teacher who decided to participate in the research, implementing the differentiated instruction intervention in one of their classes randomly allocated by the researchers, but not in the other. Whether this means that mathematics (and English and Biology) teachers are more enthusiastic for this intervention, or whether it is believed that the intervention is more suitable for those subjects remains unclear.

Student performance is measured as the average grade of the student for the subject where the student received differentiated instruction, by the end of the intervention.

Attrition analyses

In this study we initially started with 1267 students from 46 classes from eight different schools, with 20 different teachers participating with at least two classes. The 20 participating teachers made sure that all 1267 students filled out the pretest of the MSLQ. Throughout the study we lost 39 students who left school or switched classes or even tracks, ending up in classes of which the teacher did not participate in the study. The final dataset therefore consists of 1228 students.

Attrition in grade-analysis sample A total of 67 students did not have a pretest grade, which were mostly students from two classes participating the subject Dutch in upper secondary school who simply had not had any test yet at the time of the pretest measure. A total of 81 students did not have a posttest grade (of which 11 also did not have a pretest grade), mostly due to a couple of classes participating for the subject English where grades were missing. Although the 137 students that have missing pre- and/or posttest grade information are significantly different from the rest of the sample (on ability, age, grade level, level, primary school advice, teacher and school), they are roughly equally divided over treatment ($n=61$) and control group ($n=76$). Further analysis on the comparison between treatment and control group for the removed students reveal that the only significant difference between treated and untreated students can be found at the teacher (and therefore school) level. This is not surprising, given the earlier observation that pre and posttest grades were

mostly missing for a few very specific classes. This leaves a total of 1091 students for which we have both pre- and posttest grades.

When running the regressions, we lose another 15 students for whom we do not have the primary school advice. If we include CITO score as measure for ability (instead of, or in addition to primary school advice) this reduces the number of observations for which we have all data to 850. The 226 students that are then not taken into account have on average a higher pretest score and are more often female. Apart from that, there is no difference between students with and without CITO score, and there are no significant differences between treatment and control group for the remaining 850 students, except on gender (but the difference in share of girls between treatment and control has become much smaller in comparison with the larger sample of 1091 observations, in which we also found a significant difference for gender).

Attrition in MSLQ-analysis sample Of the 1228 students in our final dataset, a total of 623 students also filled out the MSLQ *posttest*. In a total of three of the eight participating schools, consisting of nine classes, the posttest of the MSLQ was not filled out by *any* of the students. Of the 605 students that did not write a posttest, 35% (210 students) came from these three schools. Note that these were also schools where no members of the research group were based, again indicating the importance of really staying on top of things when doing research in educational practices. Another 34% of the missing students on the posttest of MSLQ (202 students) are part of entire classes within the remaining five schools that also did not write the posttest. In this total of 69% of the missing posttest cases it was never individual students' (indirect) choice to not write the posttest, but the teachers (indirect) choice to not make time for it, to not find it important (enough) or to simply forget it (despite an invitation and a reminder being send out to participating teachers). Note that attrition at the level of the school and/or classroom is quite common when conducting field experiments in education (Borghans et al., 2016). Building an even better relationship with participating schools and teachers, in order to convince them of the importance for both the research and for their own learning experience in this study, might decrease the attrition rate. However, years of experience with field experiments in education have shown that there will always be *some* attrition, no matter how much time and effort you invest.

The remaining 30% of missing participation cannot be clearly identified by overall group characteristics. Overall, participation in the posttest of the MSLQ turns about to significantly dependent on the school and teacher (as explained above), but besides that on gender and ethnicity (girls and Dutch students filled out more often. Not by deliberate choice, but most likely because they are absent less often and therefore were present when the posttest was being filled out in class). When comparing the students that did and did not fill out the posttest for the MSLQ on their outcomes on the pretest of the MSLQ, we see that people that did not write the posttest score slightly, but significantly, higher on intrinsic and extrinsic goal orientation, on critical thinking and on peer learning. Furthermore, they score slightly lower on the scale of time and study environment. Note that the differences are very small, and as discussed most of the group that did not write the posttest did not individually choose to do so. This implies that the school and/or classes that dropped out have on average these higher scores on the pretest MSLQ. However, the probability of filling in the posttest does not differ between experimental and control students on any of the background or MSLQ-pretest characteristics.

Therefore, although the attrition in both the grade-sample and the MSLQ-sample is not completely random, the attrition does not affect the internal validity of our study. However, as we lose three of the eight schools in the MSLQ-sample, it does influence the external validity of the study.

In the remainder of this paper, all tables on the MSQL will therefore be based on 623 students, whereas all tables and analyses regarding student performance are based on the original 1228 students (although we did check whether the results are very different if we use the 623 students only for the student performance analysis, which is not the case).

Implementation check

About halfway the intervention, two teacher members of the research group conducted two open unstructured interviews with three to four students (from different classes and differentiation groups) and four classroom observations in two of the eight participating schools, for which a rubric was created. The schools were chosen randomly.

The classroom observations showed that the differentiation seems to take place mostly with respect to the independence of the students and the pace in which they work (implying that they get more of the tasks done). However, the goal of the lesson and the tasks are in three out of four observed classrooms the same for all the students, regardless of the group they are in. Furthermore, the instruction itself is not necessarily different, although the lower group does get extended instruction in all four lessons that are observed. In all lessons there are not that many questions from the student, and the attitude towards their schoolwork is quite expectantly. Although students work on the tasks (in their own pace) they do not take initiative to extend their own tasks when they are finished with all the provided tasks. Two of the observed teachers really come back to how the different groups worked during this lesson, whereas the other two teachers close the lesson in a general way without focusing on the different groups separately.

From the interviews with students we learned that in general the students appreciated that they could work in separate groups and either (in the case of the students in the lower group) received additional attention from the teacher that they needed or could work in their own pace and on their own topics (in the case of students in the upper group). However, they also said that they would have liked to choose the groups themselves (the students in the middle group said that they would have wanted to be in the upper group, whereas the students in the lower group said that they would have chosen the group they were in). Furthermore, the students mentioned that their preference for a group could differ by the topic of the chapter they were working on (and some teachers informed the researchers that they indeed allowed switches depending on the topic).

All in all, the intervention seems to have been implemented more or less as was originally planned, although the difference between the middle and the higher group was perhaps less clear than envisioned, and the differentiated instruction was less present than perhaps envisioned beforehand. However, the differences in learning strategies that were used to divide students over groups definitely came back in the way the lessons were organized around the three different groups.

Analysis

The data are analysed using multilevel regression, which analyzes at the student level, but in which both the teacher and the school level are taken into account as the second and the third level in our model. All student characteristics that are presented in Tables 2 and 3 are included as covariates in the model, except ability, because we have many missing observations on ability and the correlation with track advice is very high and highly significant ($r=0.73$, $p<0.000$). We therefore include gender, age, Dutch nationality, track advice, track and grade as control variables. Furthermore, we control for pretest. We include control variables to make the estimates of the experiment more precise, and to check whether the coefficient of the experiment stays stable also after including control variables.

The pretest of grades is based on the average previous grade (from earlier that academic year) on the same subject as the students participated in the experiment with, graded by the same teacher. The pretest of the MSLQ is literally the pretest of the MSLQ-questionnaire on the particular scale in that analysis. All control variables are significantly correlated to the outcome measure. Furthermore, we control for gender since performance on several subjects is different by gender. We control for age and grade because younger (older) children in their class may have skipped (repeated) a grade and it is known that this positively (negatively) relates to performance. Especially for the language subjects, Dutch nationality often influences performance. Track advice is a measure of ability, as is track itself. We checked the homogeneity of regressions slopes and have reported about those in Appendix 2.

Note that the results presented below are unstandardized. Therefore, the analysis on grades are to be interpreted based on a scale from 1 to 10 and the analyses on the MSLQ-scales are to be interpreted on a scale from 1 to 7. Since all teachers have both treatment and control classes, this will decrease the teacher effect on grades.

Table 4 Comparison experimental and control group on background characteristics

Variable	Control group		Experimental group		T-stat
	Mean	N	Mean	N	
Pretest	6.89	590	6.79	571	-0.2
Age	15.41	628	15.38	599	0.31
Gender (1=woman)	0.52	628	0.57	600	-1.72
Ability	541.26	479	541.41	481	-0.37
Dutch speaking	0.98	628	0.99	600	-1.82
					F-stat
School					2.41*
Teacher					1.4
Track advice					0.71
Grade					0.38
Grade level					0.28
Subject					1.11

* $p<0.05$; ** $p<0.01$, *** $p<0.001$

Comparability of experimental and control group

Table 4 shows the comparability of the experimental students with the control students on the observable characteristics of the students that are registered in the schools' student administrative system. We use T-test for continuous and dummy variables and ANOVAs for the categorical variables. Table 4 shows that none of the student characteristics is significantly different between the two groups. However, we do see significant differences in the share of experimental and control students between the schools, which can be explained by the fact that we do not have an even distribution of classes over schools (some schools have more participating classes than other schools).

In Table 5 we present the comparability of the students on the MSLQ of the pretest. On most scales, the experimental group does not differ from the control group, except for the scales of test anxiety and rehearsal. Although we control for the pretest scores in our analyses, the analyses on these two scales should be treated with caution.

Results

Effect of group differentiation on student performance

Table 6 shows three multilevel regression models. Model 1 presents the analysis with only the pretest as control variable. Model 2 incorporates the track advice, gender, age, ethnicity and grade year as additional control variables. Model 1 shows a significant effect of the intervention of 0.12 points on the learning results of the students. This corresponds to a small standardized effect size of 0.1 of a standard deviation. Adding the control variables in Model 2 changes the coefficient slightly to 0.11 which also shows significance at the 5% level. Lastly, Table 6 shows that the pretest has a significant coefficient in both models with

Table 5 Comparison experimental and control group on pretest

Variable	Control group		Experimental group		T-stat
	(N=305)		(N=318)		
	M	SD	M	SD	
<i>Motivation</i>					
Task value	4.7	1.06	4.8	0.95	-1.88
Self-efficacy for learning and performance	4.6	1.02	4.7	0.97	-0.43
Test anxiety	3.6	1.28	4.0	1.27	-3.27***
<i>Learning strategies</i>					
Rehearsal	4.6	1.14	4.8	1.09	-2.91**
Elaboration	4.1	0.89	4.2	0.97	-0.67
Organization	4.1	1.23	4.2	1.20	-1.57
Critical thinking	3.7	0.88	3.7	0.95	-0.60
Metacognitive self-regulation	4.3	0.73	4.4	0.75	-1.56
Time and study environment	4.7	0.76	4.8	0.84	-0.73
Effort regulation	4.9	0.93	5.0	0.90	-1.17

* $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$

Note that we underlined the learning strategies on which the division of treatment students into groups is based

Table 6 Multilevel multivariate regression of learning results

Variabele	Model 1	Model 2	Model 3
Pretest	0.670*** (0.021)	0.661*** (0.022)	0.660*** (0.022)
Experimental group	0.120** (0.045)	0.105* (0.045)	
lower group			0.171* (0.086)
average group			0.069 (0.052)
upper group			0.157* (0.074)
Gender		0.062 (0.046)	0.061 (0.046)
Age		-0.029 (0.039)	-0.031 (0.039)
Dutch nationality		0.187 (0.236)	0.177 (0.236)
Track advice prevocational education		Ref.	Ref.
Track advice higher general education		0.030 (0.071)	0.036 (0.071)
Track advice pre-university education		0.071 (0.091)	0.082 (0.092)
Track prevocational		Ref.	Ref.
Track higher general		-0.290 (0.233)	-0.284 (0.232)
Track pre-university		-0.581 (0.342)	-0.577 (0.341)
Grade		0.176* (0.079)	0.177* (0.079)
(constant)	1.961*** (0.151)	1.878** (0.616)	1.918** (0.616)
N	1091	1076	1076
Random Effects constant school	0.000	0.000	0.000
Random Effects constant teacher	0.208	0.153	0.152
Standard deviation residual	0.726	0.723	0.723
ICC school	0.000	0.000	0.000
ICC teacher	0.076	0.043	0.042

Standard errors in parentheses

* $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$

a very similar magnitude. Note that if we run Models 1 and 2 for the subsample of students for whom we also have the MSLQ-pretest, we find very similar coefficients, although not significant ($p = 0.231$ and $p = 0.114$ respectively), which is most likely due to power issues, which is not surprising given the small effect size.

In Model 3 of Table 6 we made a distinction within the treatment group between the three separated groups in class. Note that this is not a perfect measure, because, as mentioned before, teachers had some discretion to move students to a different group if they felt that

that was a better fit for that student based on their learning strategies. Model 3 shows that the effect on grade is mainly driven by the students in the lower (instruction-dependent) and upper (instruction-independent) groups, and that we do not see a significant effect on the learning results for the middle group.

Effect of group differentiation on student motivation, metacognition and learning strategies

Table 7 shows the regression results for the various models with MSLQ-scales as outcomes, where the outcome is the posttest on this MSLQ-scale, while controlling for the score on this scale in the pretest (so this boils down to the difference between pretest and posttest). Table 7 presents the regression coefficient of the experimental dummy of a multilevel multivariate regression including track advice, gender, age, ethnicity and grade year as additional control variables, similar to Model 2 in Table 6. For reasons of brevity¹ we only present the

Table 7 Multivariate regression of MSLQ variables

Scale MSLQ	Variable Experimental Group	ICC school	ICC teacher
Motivation			
Task value	0.079 (0.068)	0.193	0.000
Self-efficacy for learning and performance	-0.104 (0.077)	0.000	0.002
Test anxiety	-0.003 (0.075)	0.000	0.000
Learning strategies			
Rehearsal	-0.013 (0.076)	0.000	0.000
Elaboration	0.144* (0.067)	0.059	0.000
Organization	0.074 (0.079)	0.000	0.005
Critical thinking	0.110* (0.045)	0.000	0.000
Metacognitive self-regulation	0.085* (0.042)	0.000	0.000
Time and study environment	0.057 (0.056)	0.000	0.000
Effort regulation	0.207** (0.070)	0.000	0.000
N	626		

Standard errors in parentheses

* $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$

Note that we underlined the learning strategies on which the division of treatment students into groups is based

¹ Full regression results available upon request from the corresponding author.

coefficient of the experimental dummy in Table 7, separately for each of the MSLQ-scales. We also do not present Model 3 here, but while discussing the results we take into account which group(s) drive(s) these results.

In Table 7 it can be seen that for none of the motivation scales there is a significant effect of the intervention. However, for the learning strategies scales we do see some significant differences between the experimental and control group. The experimental group has grown significantly more on the scales of elaboration, critical thinking, metacognitive self-regulation, and effort regulation. This implies that three of these four significant scales were also represented in the set of scales that were used to divide students into the three different differentiation groups. It seems that specifically these skills were further developed by working in groups with students that have a similar learning strategy. Interestingly, where the learning outcomes were mostly affected for the instruction-dependent and the instruction-independent group of treated students, for the learning strategies outcomes we do not see a clear pattern. The effect on elaboration is driven by the instruction-dependent group, the effects on critical thinking are driven by the instruction-independent group, and the effect on effort regulation is driven by the average group. We do not see significant differences for metacognitive self-regulation, which is most likely due to power issues, given the smaller effect size.

Teacher perceptions

Teacher perceptions were measured by the teacher questionnaire. This questionnaire was filled in by 15 of the 20 teachers. All statements were answered by all teachers but only rarely was the possibility to add a comment used. We discuss the results for separate statement, and did not create scales. Only question 9, whether students had sufficient skills to work in the three groups, was complemented with comments by nine teachers. Therefore, the goal of the questionnaire to get a deeper understanding of the implementation process and to find explanations for the quantitative results was not completely reached. We propose to interpret these results with caution and to see them as exemplary in nature.

Expectancy

The first part of the questionnaire consists of the subscale of expectancy. To which extent do teachers feel that they have implemented the proposed differentiated instruction on learning strategies method in their classroom? In this subscale, most teachers scored in the upper range of the 5-point Likert scale on all items. Considering teachers' *self-efficacy and skills*, they all believed that they were effective in implementing differentiated instruction in their lessons, although they reported having little experience with this form of differentiation. Teachers who filled in the open questions commented that the differentiated instruction model was very clear and easy to use. Concerning previous experience, only three teachers reported to have experience with differentiation in this way. With respect to the training sessions, nine teachers experienced the training sessions as neutral or bad, of which five teachers commented that the trainer did not have enough expertise and did not provide any hands-on guidelines to implement the intervention. Most teachers reported that there was not enough material available to feel competent to implement this model. Most teachers (N=10) felt being in control of the learning process of all three groups. The teachers who

disagreed with this statement felt uncertain about the learning process and outcomes of the instruction independent group. This method of differentiated instruction was found feasible for only five teachers. The other teachers thought that the method demanded too much time and effort for only marginal yield. Considering adaptivity to *student characteristics*, most teachers thought students were old enough to be able to work with this model although most teachers questioned whether the students have the necessary skills to work this way. The *classroom environment* felt sufficient for five teachers to implement this model. Seven teachers explained that there was no room to have three groups seated in the regular classroom. Therefore, the instruction-independent group worked mainly without supervision which was not preferable for the teachers. Finally, considering *collegial and managerial support*, five teachers scored neutral on this question; five teachers agreed with the statement and five teachers disagreed with the statement. Answering the question whether they could rely on the support of their managers, eight teachers replied neutrally.

Value

Perceived value of the differentiated instruction method was measured through seven items. All teachers thought the method was beneficial in supporting student autonomy and all but one teacher perceived the method as a valuable method for personalized learning. Furthermore, the method was considered to stimulate students' positive attitudes towards learning. When asked whether differentiated instruction delivers some information about the student's self-regulated learning abilities, a small majority (eight teachers) confirmed and four teachers did not think they gained any information on self-regulation. Most teachers answered neutrally when asked whether differentiated instruction based on learning strategies would increase students' performance and we did not see a consistent pattern in the answers on the question whether one group would outperform other groups by using this method of differentiated instruction.

Effort and resources

In teachers' effort and resources, we found a large standard deviation on all six items as well, suggesting that there are a lot of differences between teachers. Considering preparation time, a small majority of teachers (N=9) stated that the needed preparation time in general was too large. The time needed to adjust materials to this specific differentiation method delivered mixed results with six teachers thinking it was too time consuming to adjust materials and six teachers disagreeing. Considering time in class, teachers agreed that this method did not demand more time than other instructional methods. Resources used were mainly materials developed by the teachers themselves. Eight teachers believe it is impossible to use differentiated instruction based on learning strategies without specialized materials. Combining these two aspects, resources and effort versus output, we asked the teachers about their perceptions on efficiency by answering two questions. When asking directly about their perceptions of efficiency, the majority of the teachers (N=8) thought the method was an efficient strategy to optimize learning. When asking whether the costs of using this model are higher than the benefits, only four teachers thought the costs were too high for the benefits.

All in all, based on the answers of 15 teachers we can conclude that they see the theoretical benefits of differentiated instruction on the performance and motivation of students. However, the absence of materials, the resources and effort needed, and the logistics of this operation in the classroom makes the implementation of differentiated instruction based on teaching strategies less feasible for the teachers who answered the questionnaire.

Conclusion and discussion

In this study, we analyzed the effect of group differentiation by students' learning strategies in secondary education in the Netherlands. In total around 1200 students taught by 20 different teachers in 46 classes in eight different schools participated in this study. We had a least two classes per teacher and randomized each class into either an experimental or a control group, with the constraint that each teacher had at least one control class and one experimental class. Students in an experimental class were divided into three groups, based on their learning strategies: an instruction-independent group, an average group, and an instruction-dependent group. Learning strategies were measured with a pretest based on the Motivated Strategies for Learning Questionnaire (MSLQ). In the experimental group, students were taught according to a differentiated direct instruction model, with different levels of instruction given by the teacher, depending on the group they belonged to. The control group was taught as one group, as always. The posttest consisted again of the MSLQ, of student performance measured by grades, and of a teacher questionnaire to better understand the process and underlying teacher beliefs and actions. The results showed that, when controlling for the teacher in the multilevel model, student performance is higher in classes which belonged to the experimental group, with a small standardized effect size of 0.1 of a standard deviation. Treated students score higher at the posttest of the MSLQ, but only at some of the scales that were used to divide them into groups. However, since these results are only found once we control for differences between teachers, this implies that the way the teacher implements the intervention is crucial to its effectiveness. Furthermore, the results on learning performance are driven by the instruction-dependent and instruction-independent group of students, whereas the results on the learning strategies were driven by all three groups, depending on the outcome measure.

The teacher questionnaires showed that teachers see the theoretical benefits of differentiated instruction but find the implementation of the differentiated instruction based on learning strategies less feasible. The latter is mainly based on their perception of the absence of materials, the resources and effort needed, and the logistics of this operation in the classroom. These difficulties ensured that almost half of the teachers (eight out of 20) were not inclined to use this differentiated instruction method in combination with the three groups of types of students in the future. However, as became clear in the classroom observations, they did manage to create differentiation between the groups with respect to pace and independence of the students. These findings from the teacher questionnaires, in combination with the positive effect of differentiation that we found, also have implications for teacher education programs and in-service training. As Brevik et al. (2018) also indicated, teacher education programs have an important role in teaching future teachers to effectively differentiate in class, and to make them confident, competent and knowledgeable in using differentiation in class, despite the challenges that they will undoubtedly encounter.

The hesitance of teachers to use this model can partially explain the results of this study. We only find an effect of the intervention on student performance once we control for the teacher. This implies that the teacher plays a big role in the effectiveness of the intervention, and that differences between teachers are large (similar to what was found in Feron & Schils, 2020, and what is emphasized in Hulleman & Cordray, 2009). One potential explanation for this is that there might be a large difference in the perception of teachers in their competence level of being able to use differentiation in the classroom. Teachers only received two training meetings of 3 h as a preparation for this intervention, whereas Dixon et al. (2014) show that teachers need at least 10 h of training to feel confident that they are able to differentiate. This discrepancy between the recommendation in their study and the duration of our training (6 h in total) might explain the role of the teacher in our effect.

Another potential reason for this finding is that the experiment mainly focused on the teachers. First of all, the students were not trained to work in these differentiated groups, but it was merely assumed a sufficient competence level to work in this new way (whereas it is unclear if this can be expected of students of these ages), the success of the intervention is even more dependent on the teacher, in his/her ability to guide the students properly in class in this way of working. Furthermore, although the teachers, as well as the general management of the schools, were involved in this study, the direct supervisors and colleagues of the teachers were not directly informed or involved by the researchers. This might explain why there are so large differences in the support the teachers felt from their colleagues and direct supervisors. This meant that the implementation process depended completely and solely on the teachers, which in turn might explain why the result is so teacher dependent.

Finally, based on the answers of the teacher questionnaire and specifically the open space to comment, the question arises whether the teachers were completely knowledgeable about the difference between differentiated instruction based on learning strategies and differentiated instruction based on cognitive level. From the literature and from conversations with the teachers and management, one gets the impression that when teachers differentiate, they mainly differentiate on the cognitive level. Differentiated instruction on learning strategies is relatively scarce in the educational practice. The question is whether our teachers were able to recognize the difference between these methods. Exemplary comments like: “we are already working several years with differentiated instruction” or “this student does not fit the instruction-independent group because he is very low performing in this subject”, makes us wonder whether the teachers did not switch students to other groups without reporting. In this study, we had to rely on self-reports of the teachers, and therefore we cannot be sure that students did not switch between groups. There was only one observation moment in which teachers were observed to make sure that they implemented the intervention correctly. In future research, more attention should be paid to observing the teachers on a regular basis to make sure the experiment was implemented as intended. Another option would be to also ask the students in the posttest in which group they participated during the intervention, as a cross-reference with the original assignment.

Another constraint of this research is the total sample size of this study. Although more than 1200 students participated in this research, for whom we have background information, performance data and the pretest of the MSLQ, only about 600 students filled in the MSLQ twice, posing potential power problems for the analyses where MSLQ is the outcome. It is therefore possible that other MSLQ scales would also be significantly different if we had a larger sample. However, the coefficients of the currently insignificant MSLQ scales are

very small, so even if we have a power problem, the effect is much larger for the scales of MSLQ that relate to learning strategies. Future research should take these sample size issues into account.

Furthermore, the fact that we use unstandardized tests as outcome measure may have led to biased grades. However, the tests used in our study, especially the larger and more important tests that are conducted in test weeks, are usually developed with all teachers from the section of a specific subject for a given grade level and are graded by this same group of teachers. And in the case of our study, all participating teachers from the different schools that teach the same subject have jointly developed the materials for the differentiated instruction for the treatment groups. Therefore, it is unlikely that the potential grading bias is directly related to the intervention, implying that this does not hamper our results.

Lastly, it is possible that the voluntary participation of teachers in our study influences the results. It is possible that the assessed value of cognitive performance and motivation and metacognition is higher in our sample than in a generalized population of teachers. It could be the case that the teachers that selected themselves to participate are highly motivated regarding (this type of) differentiation and thereby really try their best to make the intervention a success. However, this high level of motivation does not show at all from the teacher questionnaire that was later administered. Furthermore, since teachers taught both treatment and control classes, the outcome measures are measured at the student level, and teachers do not choose their students or vice versa, one could also argue that this teacher selection should not influence the outcome measures. In sum, we do not have enough information to know whether teacher self-selection has influenced the results, but there is no immediate reason to assume that this was the case.

Taking these (potential) constraints into account, we can conclude that differentiated instruction based on learning strategies can be a promising way to take into account the differences of high school students in today's society. In this study, we found a significant effect of differentiated instruction based on learning strategies on students' performance and parts of self-regulated learning capacities, when controlled for the role of the teacher. Future research on differentiated instruction on learning strategies should focus more on the implementation process of this method and the necessary capacities on both student and teacher level.

Appendix 1: Teacher questionnaire

The teacher questionnaire was originally formulated in Dutch and translated for this article in English. The Dutch version is available upon request from the corresponding author.

Beliefs

Expectancy

1. My knowledge of differentiated instruction is large enough to be able to implement the method.
2. The training modules have provided me with enough tools to use differentiated instruction in the lessons.
3. I have too little experience with differentiated instruction to be able to use it properly.
4. I think I have used differentiated instruction well in my class(es).
5. There is sufficient material available to develop my lessons with differentiated instruction.
6. To be able to carry out differentiated instruction properly, I can fall back on my colleagues.
7. In order to be able to carry out differentiated instruction properly, I can fall back on my supervisor.
8. My students do not have the skills to work with differentiated instruction.
9. Differentiated instruction is not suitable for the age of my students.
10. Using differentiated instruction leads to too much tensions in the classroom.
11. Differentiated instruction is not suitable for use in my subject.
12. My students enjoy working in the three groups. Specify per group.
13. The space that I have at my disposal in (and outside) the classroom satisfies my needs in my differentiation lessons.
14. I have a preference for traditional instruction over differentiated instruction.
15. As a teacher I have a grip on what the students in the different groups are doing. Specify per group.
16. Making the pupils work in the different groups is feasible for me as a teacher. Specify per group.

Value

17. Differentiated instruction is a valuable approach to being able to serve students more tailor-made.
18. Differentiated instruction ensures that students learn better.
19. Differentiated instruction on the basis of self-regulating learning capacity focuses too little on the cognitive level of the pupils.
20. Differentiated instruction mainly improves the level of pupils of 1 group. Specify per group.
21. Differentiated instruction encourages positive feelings about learning among pupils.
22. Differentiated instruction gives more autonomy to the pupils.
23. Differentiated instruction gives me more insight into the self-regulating ability of my pupils.

Cost

1. The effort it takes me to use this model is greater than the benefits.
2. It is impossible to differentiate on self-regulating ability of the pupils without specialized materials.
3. There is enough time to adjust the lesson plans so that I can differentiate.
4. Differentiated instruction is an efficient strategy to optimize learning.
5. Differentiated instruction requires too much preparation time from me as a teacher.
6. Differentiated instruction requires too much time in the classroom.

Acknowledgements This paper has benefited greatly from the contribution of Inne Vandyck and from valuable comments from Mélanie Monfrance and several anonymous reviewers. I am grateful for the contribution of the members of the research group and the participating teachers and schools.

Compliance with ethical standards

Potential conflicts of interest None.

Research involving human participants The study took place in 2015/2016 and was developed in 2014/2015. In our school ethics approval for research designs was not common back then and therefore, unfortunately, there was no ethics approval asked.

Appendix 2 Check homogeneity of regression slopes

Variable of interest	Pretest	Gender	Age	Dutch	Grade	Advice	Track
Experimental group	0.291	0.079	0.511	1.619*	0.052	0.144	0.242*
	(0.259)	(0.092)	(0.654)	(0.795)	(0.091)	(0.104)	(0.119)
Var of interest	0.683***	0.182*	-0.126**	0.472	0.290		
	(0.029)	(0.087)	-(0.044)	(0.361)	(0.175)		
Exper group * var of interest	-0.025	0.028	-0.026	-1.525	0.156		
	-(0.038)	(0.124)	-(0.042)	-(0.797)	(0.127)		
Track (advice) higher general						-0.584	0.056
						-(0.355)	(0.121)
Track (advice) pre-university						-0.200	0.290*
						-(0.359)	(0.146)
Exper group * track (advice) higher general						-0.097	-0.151
						-(0.196)	-(0.168)
Exper group * track (advice) pre-university						-0.066	-0.216
						-(0.136)	-(0.150)
N	1076	1076	1076	1076	1076	1076	1076

Standard errors in parentheses

* $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$

Informed consent All participating schools have signed a contract on their willingness to participate and their consent to share data that was necessary for the research. Parents of participating children have been informed about the study via the schools.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abrami, P. C., Poulsen, C., & Chambers, B. (2010). Teacher motivation to implement an educational innovation: Factors differentiating users and non-users of cooperative learning. *Educational Psychology*, 24(2), 201–216. <https://doi.org/10.1080/0144341032000160146>
- Artino, A. R. (2005). Review of the Motivated Strategies for Learning Questionnaire. Connecticut: University of Connecticut. Retrieved from <http://files.eric.ed.gov/fulltext/ED499083.pdf>
- Borghans, L., de Wolf, I., & Schils, S. (2016). Experimentalism in Dutch education policy. In Burns, T., & Köster, F. (Eds.), *Governing Education in a Complex World, Educational Research and Innovation*. Paris: OECD Publishing
- Blom, S., & Severiens, S. (2008). Engagement in self-regulated deep learning of successful immigrant and non-immigrant students in inner city schools. *European Journal of Psychology of Education*, 13(1), 41–58
- Brevik, L. M., Gunnulfson, A. E., & Renzulli, J. S. (2018). Student teachers' practice and experience with differentiated instruction for students with higher learning potential. *Teaching and Teacher Education*, 71, 34–45. <https://doi.org/10.1016/j.tate.2017.12.003>
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, 112(3), 579–620
- Chow, W. C., & Chapman, E. S. (2017). Construct Validation of the Motivated Strategies for Learning Questionnaire in a Singapore High School Sample. *Journal of Educational and Developmental Psychology*, 7(2), 107–123
- Coutinho, S. A. (2006). The relationship between the need for cognition, metacognition, and intellectual task performance. *Educational research and reviews*, 1(5), 162–164
- Denessen, E., & Douglas, A. S. (2015). Teacher expectations and within-classroom differentiation. In C. M. Rubie-Davies, J. M. Stephens, & P. Watson (Eds.), *The Routledge International Handbook of Social Psychology of the Classroom* (pp. 296–303). London
- 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(2), 111–127
- Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, 98(1), 37–56
- Dunn, R., & Honigfeld, A. (2013). Learning styles: What we know and what we need. *The Educational Forum*, 77(2), 225–232
- Dunn, R., & Dunn, K. (1993). *Teaching secondary students through their individual learning styles. Practical approaches for grades 7–12*. Boston: Allyn & Bacon
- Dunn, R., Dunn, K., & Perrin, J. (1994). *Teaching young children through their individual learning styles: Practical approaches for grades K-2*. Boston: Allyn & Bacon
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current directions in psychological science*, 12(3), 83–87
- Erturan Ilker, G., Arslan, Y., & Demirhan, G. (2014). A validity and reliability study of the Motivated Strategies for Learning Questionnaire. *Educational Sciences: Theory & Practice*, 14(3), 829–833
- Feiz, P., Hooman, H. A., & Kooshki, S. H. (2013). Assessing the Motivated Strategies for Learning Questionnaire (MSLQ) in Iranian students: Construct validity and reliability. *Procedia – Social and Behavioral Sciences*, 84, 1820–1825

- Feron, E., & Schils, T. (2020). A randomized field experiment using self-reflection on school behavior to help students in secondary school reach their performance potential. *Front. Psychol*, 11, 1356. <https://doi.org/10.3389/fpsyg.2020.01356>.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. 4th Edition, Sage Publications Ltd., London
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, 34(10), 906
- George, P. S. (2005). A Rationale for Differentiating Instruction in the Regular Classroom. *Theory Into Practice*, 44(3), 185–193
- Hames, E., & Baker, M. (2015). A study of the relationship between learning styles and cognitive abilities in engineering students. *European Journal of Engineering Education*, 40(2), 167–185
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, 2(1), 88–110
- Jackson, A., & Davis, G. (2000). *Turning points 2000: Educating adolescents in the 21st century. A report of the Carnegie Corporation*. New York: Teachers College Press
- Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education*, 106(1), 166–171
- Kyprianidou, M., Demetriadis, S., Tsiatsos, T., & Pombortsis, A. (2012). Group Formation Based on Learning Styles: Can It Improve Students' Teamwork? *Educational Technology Research and Development*, 60, 83–110
- Lauria, J. (2010). Differentiation through learning-style responsive strategies. *Kappa Delta Pi Record*, 47(1), 24–29
- Lovelace, M. K. (2005). Meta-Analysis of Experimental Research Based on the Dunn and Dunn Model. *The Journal of Educational Research*, 98(3), 176–183. <https://doi.org/10.3200/JOER.98.3.176-183>
- Lovelace, M. K. (2010). Meta-analysis of experimental research based on the Dunn and Dunn model. *The Journal of Educational Research*, 98(3), 176–183
- Melisse, T. (2015). Differentiated Instruction: Perceptions, Practices and Challenges of Primary School Teachers. *STAR Journal*, 4(3), 253–264 <https://europub.co.uk/articles/-A-9856>
- Midgley, C., Kaplan, A., Middleton, M., & Maehr, M. L. (1998). The development and validation of scales assessing students' achievement goal orientations. *Contemporary Educational Psychology*, 23, 113–131
- Moon, T., Tomlinson, C., & Callahan, C. (1995). *Academic diversity in the middle school: Results of a national survey of middle school administrators and teachers*. Charlottesville: National Research Center on the Gifted and Talented, University of Virginia
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In Bower, G. H. (Ed.), *The psychology of learning and motivation* (26 vol., pp. 125–173). New York: Academic Press
- Pintrich, P. R. (2000). Issues in self-regulation theory and research. *The Journal of Mind and Behavior*, 213–219
- Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning
- Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire. *Educational and Psychological Measurement*, 53(3), 801–813
- Ravitch, D. (2007). *Edspeak: A Glossary of Education Terms, Phrases, Buzzwords, and Jargon*. Alexandria, VA: Association for Supervision and Curriculum Development
- Ready, D. D., & Wright, D. L. (2011). 'Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: the role of child background and classroom context'. *American Educational Research Journal*, 48, 335–360
- 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(2), 462–501
- Roiha, A. (2014). 'Teachers' views on differentiation in content and language integrated learning (CLIL): perceptions, practices and challenges'. *Language and Education*, 28, 1–18
- Roiha, A., & Polso, J. (2021). "The 5-dimensional model: A tangible framework for differentiation.", *Practical Assessment, Research, and Evaluation*, 26, 20. <https://scholarworks.umass.edu/pare/vol26/iss1/20>
- Sapon-Shevin, M. (2000/2001). Schools fit for all. *Educational Leadership*, 58(4), 34–39
- Sarrazin, P., Tessier, D., Pelletier, L., Trouilloud, D., & Chanal, J. (2006). 'The effects of teachers' expectations about students' motivation on teachers' autonomy-supportive and controlling behaviors'. *International Journal of Sport and Exercise Psychology*, 4, 283–301

- Schraw, G., & Olafson, L. (2002). Teachers' epistemological world views and educational practices. *Issues in Education*, 8(2), 99–148
- Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, 26, 113–125
- Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting Self-Regulation in Science Education: Metacognition as Part of a Broader Perspective on Learning. *Research in Science Education*, 36(1–2), 111–139
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary educational psychology*, 19(4), 460–475
- Shah, J., & Higgins, E. T. (1997). Expectancy $\times$ value effects: Regulatory focus as determinant of magnitude and direction. *Journal of Personality and Social Psychology*, 73(3), 447–458
- Subban, P. (2006). 'Differentiated instruction: a research basis'. *International Education Journal*, 7, 935–947
- Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of educational psychology*, 82(2), 306
- Taylor, R. T. (2012). *Review of the Motivated Strategies for Learning Questionnaire (MSLQ) using reliability generalization techniques to assess scale reliability*. Dissertation, Alabama (1–182)
- Tomlinson, C. A. (2000). *Differentiation of Instruction in the Elementary Grades*. ERIC Digest. ERIC Clearinghouse on Elementary and Early Childhood 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(2/3), 119–145
- Tomlinson, C. A., & Moon, T. R. (2013). *Assessment and student success in a differentiated classroom*. ASCD publisher
- van Casteren, W., Bendig-Jacobs, J., Wartenbergh-Cras, F., van Essen, M., & Kurver, B. (2017). *Differentiëren en differentiatievaardigheden in het primair onderwijs*. Nijmegen: ResearchNed
- van den Boom, G. (2009). *MSLQ – Vragenlijst Motivatie en Studietoepak*. Open Universiteit: Heerlen
- Vernooij, K. (2009). Omgaan met verschillen nader bekeken. Wat werkt? Retrieved from: <http://www.onderwijsmaakjesamen.nl/actueel/omgaan-met-verschillen-nader-bekeken-wat-werkt/>
- Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. *Motivation and self-regulated learning: Theory, research, and applications*, 2, 297–314
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. *Handbook of self-regulation* (pp. 13–39). Academic Press
- Zimmerman, B. J., & Labuhn, A. S. (2012). *Self-regulation of learning: Process approaches to personal development*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.