Frontline Learning Research
An official journal of EARLI

Frontline Learning Research Vol. 11 No. 1 (2023) 57 - 93
ISSN 2295-3159

What do students think about differentiation and within-class achievement grouping?

Emilie J. Prast[a], Kim Stroet[a], Arnout Koornneef[a], & Tom F. Wilderjans[bcde]

[a]Educational Sciences Program Group, Institute of Education and Child Studies, Faculty of Social and Behavioral Sciences, Leiden University, Leiden, the Netherlands
[b]Methodology and Statistics Research Unit, Institute of Psychology, Faculty of Social and Behavioral Sciences, Leiden University, Leiden, the Netherlands
[c]Leids Universitair Medisch Centrum (LUMC), Leiden Institute for Brain and Cognition (LIBC), Leiden, the Netherlands
[d]Research Group of Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium
[e]Department of Clinical Psychology, Faculty of Behavioural and Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

Article received 7 April 2022/ revised 19 February 2023/ accepted 22 February 2023/ available online 23 March 2023

Abstract

Differentiation and achievement grouping are frequently implemented practices to adapt education to students' varying educational needs based on achievement level. Potential didactical and socioemotional advantages and disadvantages of these practices have been discussed in the literature. However, little is known about the perspective of students themselves. This study examined how Dutch students (N = 428) perceived differentiation and within-class homogeneous achievement grouping in primary mathematics education, with attention for potential differences between students of diverse achievement levels. Students of Grades 1, 3 and 5 completed a questionnaire about various differentiated mathematics activities and (if applicable) within-class achievement grouping. In line with the didactical perspective on differentiation, extended instruction and less difficult tasks were appreciated most by low-achieving students whereas more difficult tasks were appreciated most by high-achieving students. Students of all achievement groups had largely positive attitudes about achievement grouping and about their own achievement group. However, some differences between achievement groups were found, with less favourable results for students placed in low achievement groups. Students' responses to open-ended questions provided additional insights into the reasons behind students' evaluations of differentiation and achievement grouping. Differences between grade levels were also explored.

Keywords: Differentiation; ability grouping; student perspective; mixed methods; mathematics education.

Corresponding author: Emilie Prast, Faculty of Social and Behavioral Sciences, Institute of Education and Child Studies, Leiden University, Wassenaarseweg 52, 2333AK Leiden, The Netherlands. e.j.prast@fsw.leidenuniv.nl.
DOI: https://doi.org/10.14786/flr.v11i1.1079

1. Introduction

Many teachers strive to adapt education to their students' diverse educational needs by implementing differentiation ('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). However, differentiation is a controversial topic in the literature, particularly when it is organised by grouping students of a similar achievement level. Researchers have discussed potential didactical and socioemotional advantages and disadvantages of various types of differentiation and achievement grouping (e.g., Campbell, 2021; Francis et al., 2017; Marks, 2013; McGillicuddy & Devine, 2020; Tieso, 2003; Tomlinson et al., 2003; Van Geel et al., 2018). In this debate, students' voices have not often been heard. Given that the aim of differentiation is to adapt education to students' needs, it is important to examine whether students themselves perceive the adaptations as successful in meeting their educational and socioemotional needs. Therefore, this study investigates what students think about differentiation and within-class achievement grouping in primary mathematics education.

1.1 Differentiation and achievement grouping

Differentiation based on students' current academic achievement level (also called readiness-based or cognitive differentiation) entails two related processes: (1) monitoring students' progress to determine their current achievement level and educational needs, and (2) adapting learning goals, instruction and practice to students' current level of knowledge and skills and their corresponding educational needs (Prast et al., 2015; Roy et al., 2013). Differentiation may be convergent or divergent (Blok, 2004). In convergent differentiation, all students work towards the same goals, but the way in which students reach these goals is differentiated (e.g., with additional instruction). In divergent differentiation, students of different achievement levels work towards different learning goals.

One frequently used way to organise differentiation is to group students based on achievement level. Such groups may be homogeneous (similar achievement) or heterogeneous (mixed achievement), and within-class or between-class (see Tieso, 2003 for an overview of grouping practices). This paper focuses on within-class homogeneous grouping, that is: subgroups of students with a similar achievement level within a class which includes a broad range of achievement levels. We use the term achievement grouping rather than ability grouping since recent guidelines for grouping and differentiation do not assume that students have a fixed ability level and instead emphasise that grouping arrangements should be flexible and responsive to changes in students' educational needs (Prast et al., 2015; Tomlinson et al., 2003; van Geel et al., 2018). Achievement grouping can be used to differentiate instruction (e.g., additional instruction for subgroups with similar instructional needs) and practice (e.g., with tiered tasks for low-achieving, average-achieving and high-achieving students) (Prast et al., 2015). This paper does not only examine students' views on the grouping itself, but also on differentiated mathematics activities which may or may not take place in achievement groups.

1.2 Mathematics education in the Netherlands

Since the implementation of differentiation relies heavily on domain-specific pedagogical content knowledge (Prast et al., 2015; Van Geel et al., 2018; Vogt & Rogalla, 2009), teachers' implementation and students' perceptions of differentiation are likely to be domain-specific. To study students' perceptions of differentiation and achievement grouping in sufficient depth, this study focuses on one domain and context, namely primary mathematics education in the Netherlands. This is a relevant context, because of the increased focus on data-based decision making and, accordingly, on progress monitoring and instructional adaptations in the subject of mathematics in the Netherlands over the past decade (Prast et al., 2018; Van Geel et al., 2016, 2018; Visscher, 2015). A recent review (Prast & Hickendorff, in press) about the implementation of differentiation in primary mathematics education in the Netherlands indicated that most teachers differentiate instruction and practice based on students' achievement level at least to some extent. A mathematics lesson typically starts with a whole-class instruction. Subsequently, additional instruction is often provided to low-achieving students, whereas practice tasks are frequently differentiated at three levels.

In the lower grades, differentiation is largely convergent since students work towards the same learning goals, but from Grade 4 onwards the learning goals may also be differentiated (Expertgroep doorlopende leerlijnen, 2008). Differentiation is frequently organised using within-class achievement groups (i.e., subgroup instruction and tiered tasks), but alternatives such as individualised differentiation of the practice tasks using software are also used (Prast & Hickendorff, in press).

1.3 Perspectives on differentiation and achievement grouping

Various theoretical perspectives on differentiation and achievement grouping can be taken. In the current study, we focus on two perspectives: a didactical perspective (concerning the teaching and learning of mathematical content, with a focus on cognitive processes) and a socioemotional perspective (concerning the social and emotional processes that may be involved when differentiated activities and achievement grouping are used). We formulated our hypotheses based on these two perspectives. While other perspectives might also be taken (e.g., a sociological perspective concerning implications of differentiation and achievement grouping at a societal level), we feel that these two perspectives are most relevant in the context of the current study, because they are closely related to students' daily experiences with differentiation and achievement grouping (and therefore, relevant and understandable for students).

1.4 A didactical perspective on differentiation and achievement grouping

From a didactical perspective, the rationale for readiness-based differentiation is that adapting instruction and practice to students' current skill level enhances learning (Tomlinson et al., 2003). According to this view, learning tasks should be at a moderate difficulty level in relation to a student's current skills (Csikszentmihalyi, 1990; Murray & Arroyo, 2002). When tasks are too easy, this may result in boredom and withdrawal, while confronting a student with tasks that are too difficult may lead to frustration and anxiety (Csikszentmihalyi, 1990; Murray & Arroyo, 2002). When tasks are designed to be just within reach based on the skill level of the student, this may enhance students' motivation and achievement (Arroyo et al., 2014; Csikszentmihalyi, 1990). More generally, aptitude-treatment interaction theory predicts that students need different instructional treatments, dependent upon their aptitude (readiness for learning based on current achievement level) (Cronbach & Snow, 1977; Kalyuga, 2007). For example, direct explicit instruction may be highly effective for students with low prior knowledge but not for students with high prior knowledge (Kalyuga, 2007; Kirschner et al., 2006).

In the literature, research relating differentiation and achievement grouping to student achievement is described. For example, a meta-analysis (Deunk et al., 2018) in primary school found positive effects of interventions in which software was used to assist teachers in implementing differentiation, by continuously monitoring students' achievement level and providing (suggestions for) differentiated instruction and practice. Programmes in which differentiation was part of a broader school reform also had positive effects. Within-class grouping had no overall effect on achievement (for all students together), but there was a negative effect on the achievement of students placed in low achievement groups. However, the effects of achievement grouping were difficult to interpret, because the original studies provided little information on whether and how instruction and practice were differentiated in the achievement groups. A more recent study comparing within-class grouping and whole-class teaching in the UK (Jerrim, 2021) did not find clear evidence for effects of achievement grouping on student achievement.

From a didactical perspective, achievement grouping is merely an organisational format that can be used to implement differentiation, provided that the groups are actually used to adapt instruction to students' needs. However, achievement grouping may also have negative didactical consequences, for example if the grouping arrangements do not correspond accurately to students' current achievement level (e.g., because the groups are insufficiently flexible) or if the quality of differentiation is limited (e.g., with insufficiently challenging learning materials for lower achievement groups). More generally, if the learning goals and tasks are differentiated, low-achieving students may not get the opportunity to reach the same learning goals as their high-achieving peers (divergent differentiation; Blok, 2004; see also Hart, 1992).

1.5 A socioemotional perspective on achievement grouping

From a socioemotional perspective, achievement grouping is not just a format to organise differentiation, but an educational approach that may affect socioemotional processes within a class. First, achievement grouping may affect social comparison processes, with potential effects on students' academic self-concept. Qualitative case studies have indicated that, even when neutral names are used for the achievement groups, primary school students are largely aware of the hierarchical grouping structure, especially in the higher grades (Eder, 1983; Gripton, 2020; Marks, 2013; McGillicuddy & Devine, 2020). Campbell (2021) describes two possible mechanisms for effects of achievement grouping on academic self-concept: labelling effects and reference group effects. In the case of labelling effects, students would internalise the achievement label belonging to their achievement group, with positive effects on the self-concept of students placed in high achievement groups and negative effects on the self-concept of students placed in low achievement groups. In the case of reference group effects, students would start to compare themselves to the other students placed in their achievement group rather than to the whole class, with positive effects on the self-concept of students placed in low achievement groups and negative effects on the self-concept of students placed in high achievement groups (the big-fish-little-pond effect; Marsh, 1984, 1987). In a large-scale study about the effects of within-class grouping on students' self-concept, Campbell (2021) found more evidence for labelling effects than for reference group effects. In contrast, Jerrim (2021) found no effects of within-class achievement grouping compared to whole-class teaching on students' self-concept.

More generally, the use of achievement groups may affect the social dynamics within a class. Qualitative case studies have provided indications that placement in a high achievement group may be associated with a higher social status than placement in a low achievement group (Marks, 2013; McGillicuddy & Devine, 2020). For example, students placed in high achievement groups have been described by their peers as 'smart', 'good' or 'liked', whereas students placed in low achievement groups have been described as 'dumb', 'bad' or 'not liked' (McGillicuddy & Devine, 2020). In a study (Hargreaves et al., 2021) about peer relations of students placed in low achievement groups (including both within-class and between-class grouping systems), there was generally little evidence that troubles in peer relations were related to placement in a low achievement group. However, in some cases, students experienced feelings of exclusion that were related to their low achievement status. In a study including various types of between-class and within-class achievement grouping, students placed in low *and* high achievement groups reported achievement-related teasing (Hallam et al., 2004). Besides potential effects on peer interactions, achievement grouping might also affect teacher-student interactions. For example, teachers may implicitly or explicitly display different expectations about students placed in low versus high achievement groups (McGillicuddy & Devine, 2020; Rubie-Davies, 2014; Van den Bergh, 2018).

Taken together, this socioemotional perspective indicates that within-class achievement grouping may affect various socioemotional processes in the classroom, which may be experienced differently by students placed in low, average or high achievement groups. However, as described above, the direction of effects is not always clear and empirical research about the socioemotional aspects of within-class achievement grouping is relatively scarce.

1.6 Students' perspective on differentiation and achievement grouping

When researching differentiation it makes sense to include students' perspective, since students' motivation and engagement will be shaped by their perceptions. Little is known about students' views on differentiation and within-class achievement grouping. Since the goal of differentiation is to adapt education to students' educational needs, it is relevant to know whether students feel that their educational needs are met by the adaptations made by teachers. Besides, students can be an important source of information regarding potential socioemotional side-effects of differentiation and achievement grouping. Questions may be raised regarding the validity of student perceptions as an indicator of the "best" practice in terms of student outcomes such as achievement or motivation: we cannot expect students to oversee all implications of within-class differentiation and achievement grouping. Accordingly, the goal of this study is not to give a

complete overview of all potential effects, but to zoom in on one perspective that has not received much attention so far; that of the students themselves.

Previous research about this topic is scarce, and mostly focused on the grouping itself rather than on specific differentiation practices. First, there are studies comparing different types of grouping (e.g., between-class, within-class, whole-class teaching). Such studies have reported negative experiences with between-class achievement grouping (Boaler et al., 2000), and have presented mixed-achievement classes as a favourable alternative (Hallam et al., 2004; Tereshchenko et al., 2019). However, these studies did not focus on the differentiation practices that might be implemented within mixed-achievement classes. Second, the small-scale qualitative studies that we have described in the previous section (Eder, 1983; Gripton, 2020; Marks, 2013; McGillicuddy & Devine, 2020) provided insights into students' experiences with within-class achievement grouping, with a focus on social-emotional aspects related to the grouping. These studies did not directly ask students about their preferences regarding grouping or differentiation. Such studies are very scarce, but a third line of studies did ask students directly about their preferences regarding adaptations for students with special educational needs included in general education classrooms. Generally, students with and without special educational needs had positive attitudes towards many of these adaptations, although they wanted everybody to have the same homework (Vaughn et al., 1995; Vaughn, Schumm, Niarhos, & Daugherty, 1993; Vaughn, Schumm, Niarhos, & Gordon, 1993). These three lines of research have provided initial indications that students' experiences or preferences may differ depending on their achievement level, although the direction of effects is not always consistent across studies. For example, one study found that low-achieving students had the most positive attitudes towards mixed-achievement classes, while high-achieving students also perceived disadvantages such as a lack of challenge (Tereshchenko et al., 2019). In contrast, another study found that low-achieving students tended to prefer homogeneous grouping whereas high-achieving students tended to prefer heterogeneous grouping (Vaughn et al., 1995).

1.7 The current study: Research questions and hypotheses

The originality of the current study lies in the following. We zoomed in on students' perspective on differentiation and grouping practices *within* primary school classes. First, we made the abstract concept of differentiation more concrete by asking students about various mathematics activities and relating their evaluations to students' scores on a standardised mathematics achievement test. Second, we investigated students' opinions on within-class grouping using a mixed-methods approach combining quantitative ratings with qualitative reasons. Throughout the study, we had attention for both didactical and socioemotional considerations (rather than focusing on either), and for the perspectives of students of diverse achievement levels and grade levels.

The first aim of this study was to investigate whether students' evaluations of various mathematics activities are dependent upon their achievement level (regardless of the use of achievement grouping in their class). Within the didactical perspective on differentiation, the idea of aptitude-treatment interactions is central: students are supposed to have different educational needs depending on their current achievement level (Cronbach & Snow, 1977; Prast et al., 2015; Tomlinson et al., 2003). For example, the same activity may be appropriately challenging for some students and too difficult or too easy for other students. Note that previous research on aptitude-treatment interactions has typically focused on the outcome of student achievement, whereas students' perceived frequency, liking and learning from activities are the outcome variables in this study. Thus, the first research question was: (1) *Do different students evaluate various mathematics activities differently, depending on an interaction between the type of activity and the achievement level of the student?* In accordance with guidelines for differentiation (Prast et al., 2015), we made a distinction between general activities for all students (whole-class instruction, working at mathematics tasks independently, working at mathematics tasks together), activities intended to serve the educational needs of low-achieving students (less difficult tasks and additional instruction in a subgroup or individually), and activities intended for high-achieving students (more difficult tasks and additional instruction about these enrichment tasks, in a subgroup or individually). Based on the didactical perspective

on differentiation, we expected that students' perceptions of these activities would interact with their achievement level. First, we expected that the frequency of activities as perceived by students would be dependent on achievement level. This would be in line with previous teacher self-report and observational studies indicating that many Dutch teachers adapt instruction and practice activities to the achievement level of their students, for example by providing additional instruction to low-achieving students and more challenging tasks to high-achieving students (Prast & Hickendorff, in press). Second, we expected that students' reported liking of and learning from activities would be dependent upon students' achievement level. Based on the idea of aptitude-treatment interactions, the most probable direction of such an interaction effect would be that activities intended for low-achieving students (such as less difficult tasks) are evaluated more positively by low-achieving students whereas activities intended for high-achieving students (such as more difficult tasks) are evaluated more positively by high-achieving students. However, given the more critical views on differentiation that have also been described in the literature (e.g., Hart, 1992), as well as the innovative character of this study, it remains to be seen whether these interaction effects are indeed present and whether the effects are in the hypothesised direction.

The second aim of this study was to investigate students' perceptions of within-class achievement grouping in primary school, with attention for potential differences between students placed in low, average and high achievement groups. We were not only interested in quantitative evaluations but also in the reasons behind students' evaluations. This led to the following research questions: (2a) *How do students placed in within-class achievement groups evaluate their own achievement group and achievement grouping in general and do these evaluations differ between students placed in low, average and high achievement groups? (2b) Which reasons do students provide for their evaluations?* Based on the indications for potentially different experiences of students placed in low, average and high achievement groups in the literature reviewed above, we expected that students' evaluations would differ between achievement groups. However, given the scarce and inconsistent previous findings on student perceptions of within-class achievement grouping, we did not make specific predictions regarding the direction of those effects. Since we expected that students' reasons behind the quantitative evaluations might include socioemotional as well as didactical considerations (since grouping is typically used to differentiate tasks or instruction), we asked questions that probed both of these aspects.

2. Method

2.1 Participants and procedure

Data were collected in the fall of 2018 in the context of the research project 'Differentiation and motivation in primary mathematics education', which was approved by the local ethics committee (project number ECPW-2018/210). Fifty classes from 18 primary schools in the Netherlands participated. After obtaining active informed consent from teachers and students, data were collected by students in the final year of academic teacher training, mostly at the school where they also did a teaching internship. The schools were diverse in terms of school size, location, and pedagogical-didactical school characteristics (e.g., public schools, schools with a religious background, Montessori, etc.). We recruited one class of Grades 1, 3 and 5 (in which students are typically 6-7, 8-9, and 10-11 years old) in each school to have a spread in grade levels while retaining a substantial number of classes per grade. In multigrade classes (nine classes, 18%), only students from the grades selected for our study participated. If a class had multiple teachers (common since 67% of teachers worked part-time), the teacher who most often taught that class participated. The average class size was 23 students (range 13 – 34, including students who did not participate in the research). Teachers had an average of fifteen years of teaching experience (range 0 – 42 years). Most teachers (n = 40, 80%) were female, reflecting the general Dutch population of primary school teachers.

In the context of the overarching research project, the participating teachers were interviewed and completed a questionnaire about their differentiation and achievement grouping practices. This yielded the following background information which we provide to assist in interpreting the findings of the current study. Thirty-two teachers (64%) reported that the use of achievement groups was fully integrated in their mathematics

teaching routine. These teachers would typically start a lesson with a whole-class instruction, followed by independent practice at three difficulty levels as provided by the curricular method. Simultaneously, extended instruction would be provided to a subgroup of low-achieving students. Another fourteen teachers (28%) reported to use achievement groups partly. These teachers would for example provide extended instruction to a subgroup of students who needed it, but would either provide little differentiation in the tasks or differentiate tasks in a different way, for example using software. Four teachers (8%) did not or hardly work with achievement groups. The use of achievement groups was within-class, with one exception (in one school, mathematics was taught in separate classes for low-achieving and average-achieving students). Of the teachers using achievement groups (partly or fully), fifteen teachers (30%) indicated to create or update grouping arrangements approximately every two to six weeks based on students' scores on the end-of-chapter tests from the mathematics textbook. Another eleven teachers (22%) reported to make new grouping arrangements twice per year based on the results of a standardised mathematics achievement test. Eight teachers (16%) indicated to work with flexible groups, created per lesson or per week based on the teachers' observations, educational software or students' own view on whether they needed additional instruction. The remaining teachers created new groups 3 to 4 times per year (6 teachers, 12%), did not change the groups (1 teacher, 2%), created grouping arrangements in a different way (3 teachers, 6%) or had missing responses (2 teachers, 4%). Across the various methods of grouping, some teachers indicated that the grouping arrangements could be adapted per lesson based on students' needs, and that other sources of information such as students' daily work were also used.

Table 1

Overview of student characteristics in the full sample

Characteristic	***N***
Grade level	
Grade 1	45
Grade 3	199
Grade 5	184
Gender	
Boy	214
Girl	212
Missing	2
Achievement level on standardised test [a]	
I (highest)	97
II	57
III	60
IV	55
V (lowest)	41
Missing: test scores not available	118
Within-class achievement group [b]	
High	108
Average	93
Low	54
Not placed in single within-class achievement group during past 3 weeks	173

[a] See section 2.2.1. [b] See section 2.2.2. Students not placed in a single within-class achievement group include students who switched between achievement groups during the past three weeks, students placed in between-class achievement groups, and students from classes without achievement grouping.

For the current study, the following data were collected: a student questionnaire, student achievement group placement and student achievement on a standardised mathematics achievement test. The grouping and achievement data were collected from the teacher. The student questionnaire was administered during school hours (maximum duration: 45 minutes). In Grades 3 and 5, this questionnaire was administered to all students for whom informed consent had been obtained (n = 383). After an explanation and practice of the answering format, students completed the questionnaire independently under supervision of the research assistant. In Grade 1, the same questionnaire was administered individually due to the students' young age (typically six years). The research assistant read the questions out loud, after which the student could point to the answer (when applicable, see section 2.2.3) or say his or her answer, which was written down by the research assistant. Since this individual administration was too resource-intensive to include all students of Grade 1, we randomly selected one low-achieving, one average-achieving and one high-achieving student from the students with informed consent in each class (n = 45). Thus, the total sample consisted of 428 students with a mean age of 8 years (range 5 – 12 years). An overview of student characteristics is provided in Table 1.

2.2 Measures

2.2.1 Mathematics achievement test

Mathematics achievement was measured with the nationally administered Cito mathematics achievement tests, of which the validity and reliability have been demonstrated (Janssen, Verhelst, et al., 2005; Koerhuis & Keuning, 2011). Various grade level versions of the test are available, including a version for Kindergarteners (Janssen, Scheltens, et al., 2005; Koerhuis, 2010). Each grade level version covers multiple mathematics domains, appropriate for the grade level of the students (Janssen, Verhelst, et al., 2005; Koerhuis & Keuning, 2011). If available (administration of the test is not mandatory), the most recent test scores obtained at the end of the previous schoolyear were collected from the teacher. To ensure the comparability of scores across various grade-level versions of the test, we used the achievement level scores which reflect students' achievement level relative to a nationally representative sample: I = 80^{th} – 100^{th} percentile, II = 60^{th} – 80^{th} percentile, III = 40^{th} – 60^{th} percentile, IV = 20^{th} – 40^{th} percentile, V = 0 – 20^{th} percentile. In the analyses, these scores were recoded (and centered on the middle group) such that the highest value represents the highest achievement level (V = -2, IV = -1, III = 0, II = 1, I = 2).

2.2.2 Achievement group placement

If teachers used within-class achievement grouping, teachers were asked to indicate for each participating student in which group(s) the student had been placed during the past three weeks. The answering options were low, average, high and other (e.g., when students had switched between groups). We asked for a period of three weeks because this was long enough to experience relatively stable placement in an achievement group, but not so long that most of the students in the sample would have changed groups within that period. Since comparisons between students placed in low, average and high achievement groups might be confounded when students had switched between groups, only students placed in a single within-class achievement group during the past three weeks were included in the analyses about achievement grouping. While students' achievement group placement was generally related to their achievement on the mathematics achievement test, this correspondence was not perfect (see Appendix 1 in the supplementary materials).

2.2.3 Student questionnaire about differentiated activities and achievement grouping

The student questionnaire was developed for this study by the first and second author, based on a model for differentiation in mathematics that is frequently implemented in the Netherlands (Prast et al., 2015). Prior to large-scale administration, a small-scale pilot was conducted by administering it one-on-one to two students of Grades 1, 3 and 5 to check whether students understood the questions and the answering format. The first part of the questionnaire asked students about nine mathematics activities (based on Prast et al., 2015) representing three categories of activities: general activities (whole-class instruction, working on tasks independently, and working on tasks together), differentiated activities intended for low-achieving students (working on less difficult tasks, extended instruction in a subgroup, and individual extended instruction) and

differentiated activities intended for high-achieving students (working on enrichment tasks, subgroup instruction about enrichment tasks, and individual instruction about enrichment tasks). Based on the teacher interview, the names of the activities as used in the students' own class were used (e.g., if enrichment tasks were called "mathematics tigers", that term would be used rather than "more difficult tasks"). For each of the nine activities, students were asked how often they were engaged in this activity, how much they liked this activity, and how much they learned from it (for a total of 27 questions). The answering format was a five-point scale represented by dots as shown in Figure 1 (adapted from Park et al., 2016). Students had to indicate the dot that corresponded to their answer: the smallest dot corresponded to the smallest magnitude (e.g., never receiving whole-class instruction) and the largest dot corresponded to the largest magnitude (e.g., receiving whole-class instruction every lesson).

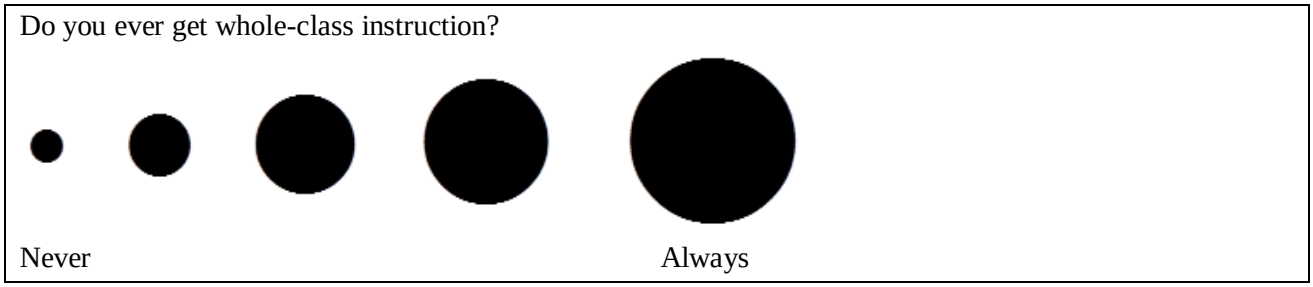

Figure 1. Sample item with answering format adapted from Park et al. (2016).

The second part of the questionnaire (consisting of 4 closed-ended and 5 open-ended questions) was only administered to students whose teachers used achievement groups. Students were asked how much they liked to be in their achievement group (answered on the dot-Likert scale described above), what they liked (open-ended) and did not like (open-ended) about being in that achievement group, how much they learned from being in that achievement group (dot-Likert scale), why they learned much or little in that achievement group (open-ended), whether and why they would rather be in a different achievement group (yes or no followed by an open-ended explanation), and whether and why they would prefer a system without achievement groups (yes or no followed by an open-ended explanation).

Note that there is no strict separation between the differentiated activities (part 1) and achievement grouping (part 2): in many classes, within-class achievement grouping was used to organise the differentiated activities. However, the first part focused on differentiated activities, regardless of whether achievement groups were used in that class, while the second part focused on students' perceptions of achievement grouping (only if relatively fixed achievement groups were used in their class).

2.2.4 Analyses

Data were analysed in two parts, corresponding to the questionnaire and our research questions. Part 1 focused on students' ratings of the various activities, relative to their achievement level as measured with the Cito mathematics achievement test. Thus, students from classes without relatively fixed achievement groups were also included in these analyses, given that differentiated activities could also be organised in different ways, whereas students for whom Cito test scores were not available were excluded from these analyses. Part 2 focused on students' perceptions regarding achievement grouping and therefore included only students who had been placed in a single within-class achievement group. Data were analysed in R with multilevel models to take into account the nesting of students within classes and schools. To enhance readability, we focus on the most important steps here while additional statistical details are provided in the supplementary materials (see Appendix 2).

To answer the first research question (*Do different students evaluate various mathematics activities differently, depending on an interaction between the type of activity and the achievement level of the student?*), we analysed whether there was a significant interaction effect between the type of activity (e.g.

whole-class instruction, independent work, etc.) and students' achievement level in predicting students' self-reported frequency of being engaged with the activities, students' liking of the activities and students' perceived learning from the activities. For each of these three outcome variables separately, we estimated four-level regression models with activity ratings (e.g., liking of whole-class instruction, independent work, etc.) nested in students (i.e., repeated measures), who were nested in classes nested within schools. To evaluate the significance of the interaction effect, we compared the fit of a full model including main effects of activity and achievement as well as an interaction between these variables to the fit of a reduced model without the interaction effect using a Likelihood Ration Test (LRT). A significant LRT indicates that the full model fits the data significantly better. Significant interaction effects were followed up with post-hoc tests evaluating the effect of achievement on the ratings of each activity. Finally, potential interactions with grade level were explored (see section 3.2.1). As described in the introduction, we expected that the interaction effects would be significant.

To answer the second research question, we performed two types of analyses: quantitative analyses to answer question 2a (*How do students placed in within-class achievement groups evaluate their own achievement group and achievement grouping in general and do these evaluations differ between students placed in low, average and high achievement groups?*), and qualitative analyses to answer question 2b (*Which reasons do students provide for their evaluations?*). The closed-ended questions were analysed using multilevel models, herewith taking the clustering of students within classes and schools into account. Achievement group was specified as a predictor of the outcome variable of interest: liking of achievement group, learning from achievement group, wanting to be in a different achievement group or preferring to work without achievement groups. As the latter two outcomes are dichotomous, a multilevel version of logistic regression was used. Likelihood ratio tests were used to determine whether this model fitted the data significantly better than a model without achievement group as a predictor. As described in the introduction, we hypothesised that there would be differences between students placed in low, average and high achievement groups, but did not have specific hypotheses regarding the direction of effects.

The analyses of the open-ended questions were exploratory and intended to give more meaning to the quantitative results. An inductive approach, in which students' answers rather than theoretical expectations formed the starting point, was taken (Linneberg & Korsgaard, 2019). Based on an initial review of students' answers, the first author developed various lower-order codes (about 10 – 20 codes per question) which specified (aspects of) answers that were given by multiple students. Since several codes were related to similar themes, these lower-order codes were then classified as belonging to one of the following higher-order themes, which were largely similar across questions: (1) answers about independent work and its difficulty level (2) answers about (social interactions within) the achievement group (3) answers related to instruction and the teacher (4) answers about learning and understanding and (5) general and other answers (mostly unspecific, e.g. "I just like it"). The coding scheme which was thus developed by the first author was discussed with the second author and revised accordingly. A random sample of 50 cases (21% of the 238 cases with achievement grouping data) was coded by both authors to determine interrater agreement. Cohen's kappa for the lower-order codes ranged between .70 and .86 for the various questions, and percentage agreement between 74.0% and 87.0%, indicating fair to good interrater agreement. The most frequent reason for non-agreement was that one of the authors had coded a statement with an unspecific code (i.e., general or other), whereas the other author had coded it with a specific code. After reviewing the cases of non-agreement, the second author understood the choices the first author had made and mostly agreed with them. The full sample was coded by the first author.

3. Results

3.1 Part 1: students' perceptions of differentiated activities

The analyses for part 1 included 310 students who provided data on both the questionnaire and the achievement test (for some analyses, *n* is slightly smaller due to missing data on single items in the questionnaire). Students' achievement test scores were distributed as follows: I (highest achievement) = 97

students, II = 57 students, III = 60 students, IV = 55 students, V = 41 students. In accordance with our sampling procedure, most students were in Grade 3 (n = 139) and 5 (n = 150). Only n = 21 students of Grade 1 were included in these analyses, since achievement data of the previous year (when they were still in Kindergarten) were not available for many students (see also section 3.1.1). Descriptive statistics and results from the model building stage can be found in the supplementary materials (Appendix 2).

In line with our hypothesis, there were significant interaction effects of activity by achievement for all outcome variables. That is, the full models including the interaction term had a significantly better fit than reduced models without this interaction term and this was the case for students' ratings of the frequency (χ^2 (8) = 182.36, p <.001), liking (χ^2 (8) = 161.75, p <.001), and amount of learning from these activities (χ^2 (8) = 95.74, p <.001). These interaction effects are visualised in Figure 2, which displays the estimated means based on the full models for students' self-reported frequency (left column), liking (middle column), and learning (right column) of activities, split by achievement level (for visual clarity, the achievement levels II and IV are not displayed, but these follow the same linear regression). The figure shows that some activities were rated more highly by low-achieving students compared to high-achieving students, whereas this pattern was reversed for other activities. For example, easier tasks (printed in red) were rated more highly by low-achieving students (red dots), whereas enrichment tasks (printed in blue) were rated more highly by high-achieving students (blue squares). In the following paragraphs, these interaction effects are interpreted further based on post-hoc tests evaluating the significance of the effect of students' achievement level on students' ratings for each activity.

First, we consider students' reported frequency of being engaged in the activities. The general activities of whole-class instruction and working together were reported equally frequently by students of all achievement levels. However, high-achieving students reported to work independently more often (although this difference seems small in the figure, it was significant). As hypothesised, the activities intended for low-achieving students (less difficult tasks, extended instruction in a subgroup, and individual extended instruction) were reported more frequently by low-achieving students. Regarding the activities for high-achieving students, high-achieving students reported to work on enrichment tasks more frequently, as hypothesised. However, high-achieving students did not report to receive more instruction about enrichment tasks, either in a subgroup or individually: these activities were generally reported infrequently, regardless of the achievement level of the students.

Second, we consider students' liking of the activities. Regarding the general activities, whole-class instruction was liked somewhat more by low-achieving students, independent work was liked somewhat more by high-achieving students, whereas working together was appreciated equally by students of all achievement levels. As expected, the activities intended for low-achieving students (less difficult tasks, extended instruction in a subgroup, and individual extended instruction) were liked more by low-achieving students. Regarding the activities for high-achieving students, enrichment tasks were liked more by high-achieving students, in line with our hypothesis. However, students' liking of instruction about enrichment tasks (in a subgroup or individually) did not depend on students' achievement level. This should be viewed in light of the relatively low reported frequency of instruction about enrichment tasks.

Third, we consider students' reported amount of learning from the activities. Students' reported learning from the general activities (whole-class instruction, independent work and working together) was not dependent upon their achievement level. As expected, low-achieving students reported to learn more from the activities intended for low-achieving students (less difficult tasks, extended instruction in a subgroup, and individual extended instruction) compared to high achieving students. Note, however, that even the lowest-achieving students rated working on easier tasks lower than the general activity of working independently (as can be seen in Figure 2). Similar to the results regarding frequency and liking, high-achieving students did report to learn more from enrichment tasks, but did not report to learn more from instruction about enrichment tasks compared to students of other achievement levels.

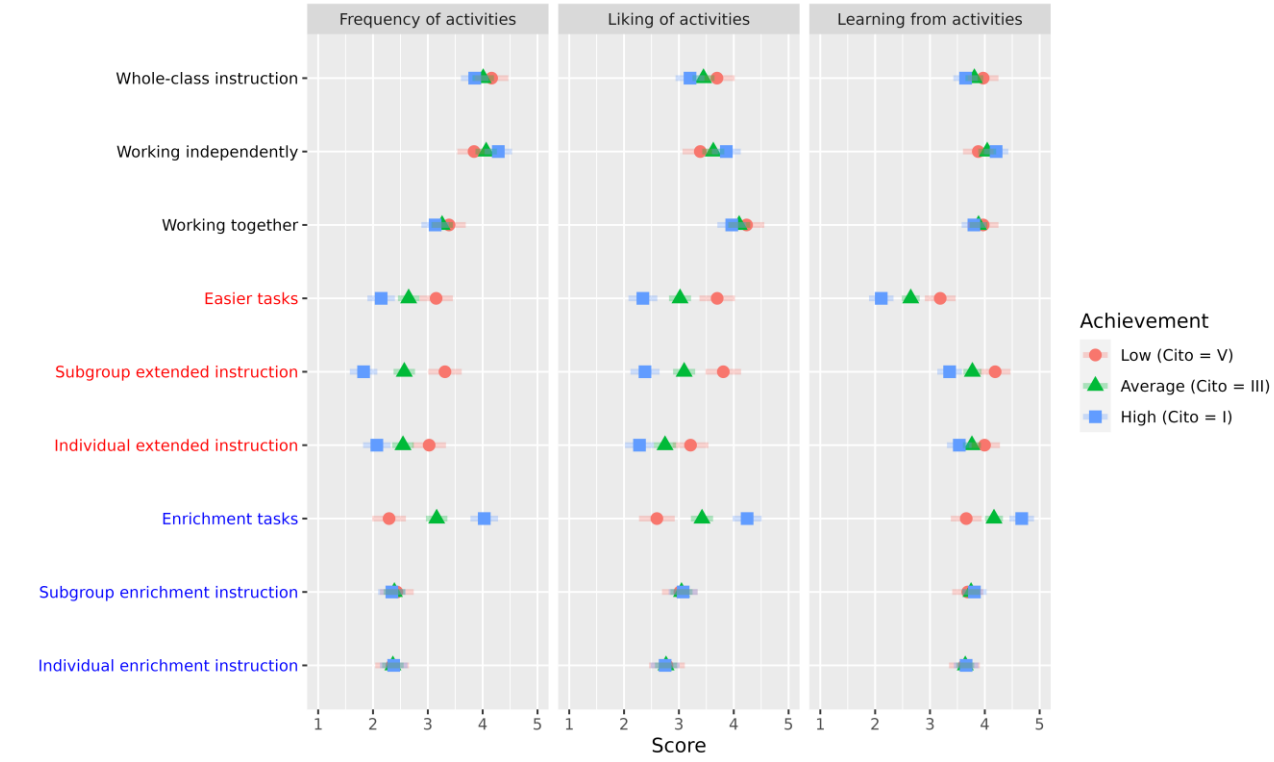

Figure 2. Estimated means of student-reported frequency, liking and learning from activities split by achievement level based on the final multilevel models. General activities are printed in black, activities intended for low-achieving students are printed in red, and activities intended for high-achieving students are printed in blue. Error bars represent 95% confidence intervals.

3.1.1 Differences between grade levels in students' perceptions of differentiated activities

We explored whether the above results varied between grade levels, by adding grade level as a variable to the analyses and examining whether there were three-way interactions of activity by achievement level by grade level. Again, the significance of the interactions was determined using Likelihood ratio tests comparing the model with three-way interaction to the model without three-way interaction. The three-way interaction was significant for students' reported frequency of activities (χ^2 (16) = 65.20, p <.001) and learning from activities (χ^2 (16) = 39.55, p = 0.001), but not significant for liking of activities (χ^2 (16) = 25.473, p = 0.062). Follow-up analyses split by grade level indicated that the activity by achievement level interaction was not significant in any of the Grade 1 analyses. However, this result should be interpreted with caution given the small sample size in Grade 1.[1] Within Grades 3 and 5, all activity by achievement level interactions were significant.

[1] We repeated the analyses with teachers' estimation of students' achievement level rather than the standardised achievement test as an independent variable, thereby increasing the sample size to n = 40. This yielded similar results, namely no significant activity by achievement level interaction effects in Grade 1.

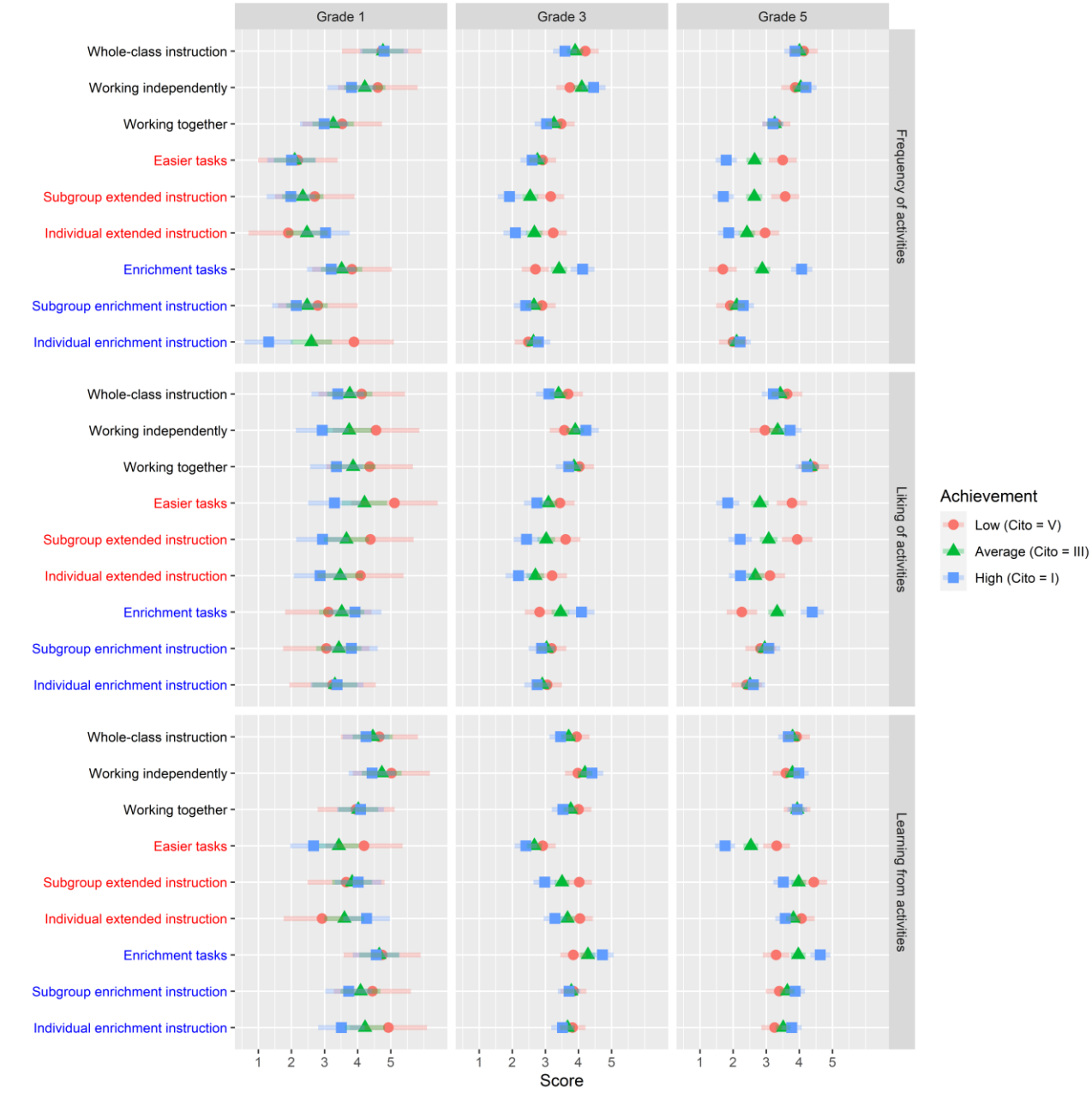

Figure 3. Estimated means of student-reported frequency, liking and learning from activities split by achievement level and grade level based on the final multilevel models. General activities are printed in black, activities intended for low-achieving students are printed in red, and activities intended for high-achieving students are printed in blue. Error bars represent 95% confidence intervals.

These similarities and differences between grade levels are illustrated in Figure 3. In the left column representing Grade 1, students' ratings of the frequency (upper row), liking (middle row), and learning (lower row) of activities have broad confidence intervals which mostly overlap between students of diverse achievement levels. This illustrates that there were no significant differences between students of diverse achievement levels, although the pattern of ratings for liking does suggest systematic differences in the expected direction. In Grades 3 (middle column) and 5 (right column), the pattern of effects was similar to the overall analyses, and the differences between low-achieving and high-achieving students tended to become more pronounced between Grades 3 and 5. For example, regarding enrichment tasks, it can be seen

that the pattern of higher ratings by higher-achieving students was already present in Grade 3 (for frequency, liking and learning of the activities), but these differences became more pronounced in Grade 5, as indicated by the larger distance between the scores of low-achieving students and high-achieving students. For easier tasks and extended instruction in a subgroup, the differences between low-achieving students and high-achieving students also seemed to increase (partially) between Grade 3 and Grade 5. However, for individual extended instruction, the difference between achievement levels did not seem to increase between Grade 3 and Grade 5, nor for the remaining activities (general activities and instruction about enrichment tasks), for which no (big) differences between achievement levels were present in the total sample. Summing up, students' ratings of activities tended to become more strongly related to students' achievement level in higher grades.

3.2 Part 2: students' perceptions of achievement grouping

The analyses of Part 2 included 240 students who had been placed in the same within-class achievement group during the past three weeks (low: n = 52, average: n = 87, high: n = 101). For an overview of all codes that resulted from the qualitative analyses and their frequency, see Appendix 3 in the supplementary materials.

3.2.1 Students' liking of their achievement group

First, we asked students how much they liked to be in their achievement group. The Likelihood ratio test indicated that the model with achievement group as a predictor of liking did not fit the data significantly better than the reduced model without achievement group as a predictor (χ^2 (2) = 5.21, p =.074). The raw means indicated that students placed in low (M = 3.94, SD = 1.24), average (M = 4.16, SD = 1.26) and high achievement groups (M = 4.35, SD = 0.98) generally liked to be in their achievement group. While the raw means increased with achievement level, the Likelihood ratio test indicated that these differences were not significant.

Most responses to the open-ended question what students liked about their achievement group were related to the higher-order theme of independent work and its difficulty level (see Table A8). While comments about this theme were made frequently by students from all achievement groups, the content of the answers differed between achievement groups. Many students from high achievement groups mentioned that they liked challenges and difficult tasks ("Because if I have this work, it's a challenge"). In contrast, students from low achievement groups tended to appreciate that tasks were not too difficult, and some also mentioned that they liked to have fewer tasks, enabling them to finish their work. Students from average achievement groups mentioned that the difficulty level was appropriate (not too difficult and not too easy) and sometimes explicitly mentioned that "it matches my level". A second frequently mentioned higher-order theme comprised comments related to the achievement group itself, including its members and the social interactions between them. Across achievement groups, students made positive comments about the members of the group ("The children in this group are kind") and about being able to help each other ("Because if you don't know something the other children can help you"). A few students from average and high achievement groups made explicit comments about liking to be in a higher group: "Then I think sometimes that I am a bit smarter together with the other children and that gives me confidence". The remaining answers were related to the higher-order themes of learning and understanding (e.g., "Then you learn more of it"), to instruction and the teacher (e.g., "That you get more help"), or belonged to the category of general and other answers (e.g., "I just like it").

Upon the question what students did not like about their achievement group, fewer than half of the students mentioned a specific aspect that they did not like (see Table A9). In fact, many students replied that they liked everything, although this was mentioned somewhat less frequently by students from low achievement groups. Most of the specific answers were related to the higher-order themes of independent work and its difficulty and the achievement group. Negative aspects mentioned by students placed in low achievement groups included too easy tasks, boredom and wanting to be in a higher group. Negative aspects mentioned by

students placed in high achievement groups included needing to work hard or fast, distraction (by other students but also by the teacher explaining to another group), difficult work, and stress. For students placed in average achievement groups, some of the answers resembled those of students in low achievement groups (e.g., too easy, wanting to be in a higher group) whereas others were similar to those of students in high achievement groups (e.g., too difficult, needing to work too hard). Relatively few students made negative comments related to the higher-order themes of learning and understanding (e.g., "I don't learn so much and I want to get better") and instruction and the teacher (e.g., "You get additional explanation when you understand it already").

3.2.2 Students' learning in their achievement group

Second, we asked students how much they learned in their achievement group. The Likelihood ratio test indicated a significant effect of achievement group (χ^2 (2) = 7.17, p =.028), indicating that students' perceived amount of learning differed between achievement groups. Students from high achievement groups (M = 4.40, SD = 0.96) perceived to learn more than students from average (M = 4.21, SD = 1.03) and low (M = 3.98, SD = 1.13) achievement groups. Note, however, that all these means are relatively high on a five-point scale.

Students' responses to the question why they learned much or little in that achievement group were frequently related to the higher-order theme of independent work and its difficulty (see Table A10). This theme was most prominent for students from high achievement groups, who mentioned frequently that they learned much because of the higher difficulty level of the tasks: "You learn much because you also get more difficult sums". When students from low and average achievement groups referred to this theme, their answers were more mixed: some mentioned an appropriate difficulty level as a reason for learning much, whereas others indicated that they did not learn so much because of an inappropriate difficulty level (too easy or too difficult). The question why students learned much or little also provoked relatively many comments related to the higher-order theme of learning and understanding. Students from average and high achievement groups tended to explain why they learned or understood more ("You learn new goals every time"), whereas students placed in low achievement groups referred to learning and understanding both positively ("I learn much because I understand it better now"), and negatively ("Because I don't understand it at all"). The higher-order theme of instruction and the teacher was mentioned by students across achievement groups, mostly in a positive way ("Because the teacher gives you more explanations and does more sums with you"). Some students also provided answers related to the higher-order theme of the achievement group, which were similar to the answers described as reasons for liking or not liking the achievement group. Again, a substantial proportion (about one-third) of the comments was classified as general or other ("I learn much because I learn much from it").

3.2.3 Students' preference for an achievement group

Third, we asked students whether they would prefer to be in a different achievement group. The Likelihood ratio test indicated a significant main effect of achievement group (χ^2 (2) = 24.68, p <.001). As can be seen in Table 2, about half of the students currently placed in a low achievement group would prefer to be in a different achievement group, compared to only 11% of students currently placed in a high achievement group. For students from low achievement groups, reasons for wanting to stay in the same achievement group included appropriate difficulty in the current group), and positive comments about the group members and interaction in the current group (see Table A11). In contrast, other students placed in low achievement groups wanted to move to another group because they wanted more difficult tasks, enrichment tasks, or more challenge, because they thought they would learn more or get better at mathematics in another group, because of the members of the desired group or for the sake of being in a higher group. For students placed in high achievement groups, reasons for wanting to stay in the same group were mainly related to appreciating the current difficulty level of the tasks and specifically enrichment tasks ("Enrichment tasks are fun, so I want to keep those"), although a few students wanted to move to a different group because they thought that the current material was too difficult. For students from average achievement groups, comments

related to the higher-order theme of tasks and their difficulty level were among the most frequent reasons (besides general comments) to want to stay in the same group, but also to switch achievement groups.

Table 2

Students' preference for being in another achievement group

Current achievement group	Preference for other group	Wish to stay in current group
Low (n = 52)	25 (48.1%)	27 (51.9%)
Average (n = 84)	30 (35.7%)	54 (64.3%)
High (n = 97)	11 (11.3%)	86 (88.7%)
Total (N = 233)	66 (28.3%)	167 (71.7%)

3.2.4 Students' preference for working with or without achievement groups

Finally, we asked students whether they would prefer to work without achievement groups. The Likelihood ratio test indicated no significant main effect of achievement group (χ^2 (2) = 1.25, p = 0.535). Across achievement groups, about 75% of students wanted to retain the achievement groups (see Table 3). The most frequent reasons for wanting to retain achievement groups included general positive comments about the current grouping system, as well as between-student differences or appropriate difficulty in the current system (see Table A12). Some students explicitly described differences between students as an argument for differentiation: "Well, if some students find it difficult and others find it hard and everybody does the same, I just don't think it's handy". Reasons for preferring to work without groups included the opportunity to learn more in a system without achievement groups and general negative comments about achievement groups. A few students explicitly mentioned equality, stating that they would like everybody to get the same tasks or to be equal. While all of these reasons for and against achievement grouping were mentioned by students of all achievement groups, the general tendency was that students from high achievement groups relatively frequently mentioned an appropriate difficulty level and the opportunity to learn more as reasons for retaining the grouping system.

Table 3

Students' preference for working with or without achievement groups

Achievement group	Without achievement groups	With achievement groups
Low (n = 51)	15 (29.4%)	36 (70.6%)
Average (n = 84)	18 (21.4%)	66 (87.6%)
High (n = 99)	22 (22.2%)	77 (77.8%)
Total (N = 234)	55 (23.5%)	179 (76.5%)

3.2.5 Differences between grade levels in students' perceptions of achievement grouping

To explore potential differences between grade levels in students' answers to the closed-ended questions, interactions with grade level were added to the analyses. None of these interactions were significant, indicating that the results were similar across grade levels. An extensive analysis of between-grade level differences in the open-ended answers is beyond the scope of this paper. We did explore the relative frequency of the answering categories across grade levels and found that students in higher grades tended to give relatively more specific answers (i.e., fewer general and other answers) than students in lower grades. In addition, relatively more of the answers of students in higher grades tended to be related to tasks and difficulty. In students' explanations of why they would prefer to work with or without groups, most of the answers referring to either equality as an argument for no grouping or between-student differences as an argument for grouping were made by students from the highest grade.

4. Discussion

While potential didactical and socioemotional advantages and disadvantages of differentiation and achievement grouping have been discussed in the literature, few studies have asked the opinion of students themselves. The current study extends the literature by exploring students' perspective on differentiated activities and within-class achievement grouping, with attention for potential differences between students of diverse achievement levels.

4.1 Students' perceptions of differentiated activities

Our first research question was whether different students evaluate various types of mathematics activities differently, depending on their achievement level. As hypothesised, there were significant interactions between the type of activity and students' achievement level for all three outcome variables: perceived frequency, liking and learning of the activities. In line with guidelines for differentiation and with previous studies in which teachers reported their use of differentiation strategies (Prast et al., 2015; Roy et al., 2013; Van Geel et al., 2018), low-achieving students perceived to receive extended instruction and less difficult tasks more frequently whereas high-achieving students worked at more difficult tasks more frequently. The infrequent occurrence of specific instruction for high-achieving students is not recommended, but also corresponds with previous findings (Inspectorate of Education, 2019; Prast & Hickendorff, in press). Regarding students' liking and learning of the various activities, we found that activities intended for low-achieving students such as less difficult tasks and extended instruction were rated more highly by low-achieving students, whereas more difficult tasks were rated more highly by high-achieving students. These are examples of perceived aptitude-treatment interactions (Cronbach & Snow, 1977; Kalyuga, 2007). However, the following observations should be kept in mind. First, scores for general activities such as whole-class instruction were also high across achievement groups. Second, students' liking and learning from activities seemed to be related to students' reported frequency of engaging in these activities. This might imply that students simply like, and perceive to learn from, activities to which they are used. Nevertheless, if students' experiences with activities would have been strongly negative, it seems unlikely that a higher frequency of an unpleasant activity would increase students' liking of that activity. Third, students generally reported to learn less from less difficult tasks, although this was less pronounced for low-achieving students than for high-achieving students. This is related to the issue of convergent versus divergent differentiation (Blok, 2004). In the higher grades of primary school, the tasks in the lowest tier of many mathematics textbooks lead towards lower end-of-school learning goals than the tasks in the highest tier (Expertgroep doorlopende leerlijnen, 2008). Thus, it may be true that students in the higher grades learn less from less difficult tasks in the sense of covering less content (regardless of the degree of understanding of that content). Finally, while these results are largely in line with our hypotheses based on the didactical perspective, students' ratings of liking and learning from activities may also have been influenced by socioemotional factors. These are discussed in the following section.

4.2 Students' perceptions of achievement grouping

Our second research question was how students of diverse achievement levels evaluate their own achievement group and achievement grouping in general. Our results provide partial support for the hypothesis that students' perceptions of achievement grouping would differ between students placed in low, average and high achievement groups. Generally, the average scores for liking and learning from one's own achievement group were quite high. Students' liking of their own achievement group did not differ significantly across groups, but students' perceived degree of learning did vary between achievement groups, with lower scores for students placed in low achievement groups. Overall, about 70% of the students were satisfied with their achievement group placement, which is more than has been reported in previous studies about between-class grouping (Boaler et al., 2000; Hallam et al., 2004). However, this question revealed the most pronounced differences between achievement groups: around 50% of students placed in low achievement groups would prefer to be in a different group, compared to only 10% of students placed in high achievement groups. Nevertheless, around 75% of students across achievement groups wanted to retain the grouping system (although this could also reflect a general desire for things to stay as they are; Hallam et al.

(2004) also found that most students did not want to change anything about the grouping practices in their school). Taken together, these results suggest quite positive attitudes towards grouping in general, but somewhat less positive experiences with placement in a low achievement group compared to a high achievement group.

In addition to these quantitative ratings, we investigated which reasons students provided for their evaluations. As expected, students' answers to the open-ended questions included socioemotional as well as didactical considerations, although didactical considerations seemed to be more prominent. Students clearly evaluated the use of achievement grouping in relation to the use of differentiated activities. In line with the didactical perspective on differentiation, many students mentioned didactical advantages including the appropriate amount and difficulty level of independent work (mentioned by students of all achievement groups), challenge (mainly mentioned by students placed in high achievement groups) and the possibility to get additional instruction and to understand the material better (mainly mentioned by students placed in low and average achievement groups). However, some students also mentioned didactical disadvantages. Some students from low achievement groups perceived the work as too easy, did not appreciate additional instruction, wanted more challenge or thought that they would learn more in a higher group. In contrast, some students from high achievement groups felt that the material was too difficult or that they needed to work too fast, which was sometimes stressful. Some students from average achievement groups made similar comments in both directions (too difficult or too easy). Ideally, differentiation should ensure that the tasks and instruction are appropriately challenging for the students in each achievement group (Prast et al., 2015). Students' answers indicate that, while many students perceived the difficulty level as appropriate, other students did not. Moreover, challenge seemed to be viewed by many students as something belonging exclusively to the high achievement group. Enrichment tasks were highly valued by students from high achievement groups, but some students from low and average achieving groups also expressed the desire to work on enrichment tasks. Compared to previous studies on between-class achievement grouping (Boaler et al., 2000; Hallam et al., 2004), the students in our sample generally seemed to be more positive about the didactical advantages of achievement grouping, but some of the perceived disadvantages resembled those mentioned in previous studies (e.g., a lack of challenge in low achievement groups, needing to work too fast in high achievement groups).

As expected, students' answers also included socioemotional considerations, but these were not always related to the socioemotional perspective on achievement grouping as described in the introduction (i.e., based on social comparisons of achievement level). Many students made comments about their achievement group that did not seem to be related directly to the achievement level of that group, such as (dis-)liking the members of the group or having positive or negative interactions within the group. This importance of peer interactions in general in students' perceptions of schooling echoes previous findings by Hargreaves et al. (2021). Partly in line with previous studies (Eder, 1983; Marks, 2013; McGillicuddy & Devine, 2020), we found some indications for social comparisons based on achievement group placement. A few students mentioned that they liked to know their own level, or the level of the other students. Students occasionally mentioned the fact that their group was low (also: "bad") or high (also: "the best" or "smart") as a negative or positive aspect of being in that group. While such comments were relatively infrequent, they support the idea that within-class achievement groups may strengthen social comparison processes by making students more aware of whose achievement is low, average or high compared to the class average (i.e., labelling effects; Campbell, 2021). Based on students' spontaneous answers to our open-ended questions, explicit teasing or stigmatisation based on achievement group placement did not seem to play a major role in the current sample. Of course, these findings do not exclude the possibility of implicit stigmatisation or social status associated with achievement group placement (see Marks, 2013; McGillicuddy & Devine, 2020; Van den Bergh, 2018). Taken together, students' answers provide some support for potential socioemotional side-effects of within-class achievement grouping, including social comparisons of achievement level, but do not indicate pervasive negative social effects of placement in a low achievement group or of achievement grouping in general. This might be partly explained by differences between countries in the way in which

achievement grouping is implemented. Since the achievement groups in this study were typically used only part of the time (besides whole-class activities) and were relatively flexible, it could be that this reduced the potential negative socioemotional effects of achievement grouping (Education Endowment Foundation, 2018). If grouping arrangements are sufficiently flexible to respond to students' current achievement level and corresponding educational needs, as recommended (Prast et al., 2015), students might evaluate them more positively than when students perceive to be stuck in a (low) achievement group. Note, however, that the degree of flexibility of the grouping arrangements differed substantially between teachers in the current sample (see section 2.1).

4.3 Differences between grade levels

We also explored whether students' views on differentiation and achievement grouping differed between grade levels. We emphasise that our findings in Grade 1 should be viewed as exploratory, given the small sample size. By and large, there seemed to be a trend towards more pronounced opinions in the higher grades. Students' reported liking and learning from activities were more strongly related to student achievement level in higher grades. The quantitative ratings of achievement grouping were similar across grades, but students in higher grades gave relatively more specific answers to the open-ended questions. This may have several reasons. First, due to maturation, older students may have been better able to express their opinions. This is likely to have affected the open-ended questions more strongly than the closed-ended questions. Second, older students may have developed more pronounced opinions about differentiation due to more experience with differentiation. Third, through socialisation, older students may have endorsed the values of an educational system which assumes that lessons should be adapted to between-student differences in achievement level (Raveaud, 2005). In future research, it would be interesting to follow a group of students longitudinally from Grade 1 onwards to examine how students' views on differentiation and achievement grouping develop over time.

4.4 Limitations, conclusions and implications

Students' perceptions of differentiation and achievement grouping were central to this study. We do not claim that students always know what is best for them in terms of learning outcomes (see Kirschner & Van Merriënboer, 2013). Nevertheless, we feel that it is important to consider what students themselves think about the degree to which differentiation is meeting their educational needs, even if it would only be to explain teachers' choices better to students if students would fundamentally disagree with the approach taken. This did not seem to be the case: by and large, students had quite positive attitudes towards differentiation.

With a self-report questionnaire, there is always a risk of socially desirable answers. However, our general impression is that students responded quite frankly, maybe also due to their young age (e.g., "I don't understand a shit of it"). Our method of data collection offered several advantages. By asking students to quantitively rate specific mathematics activities (and relating this to students' achievement level in the analyses), we could investigate the complex construct of differentiation in a way that was easy to understand for students as well as relatively quick and standardised. This enabled data collection on a larger scale and, therefore, provided more opportunities to quantify and generalise differences (or similarities) between students of diverse achievement levels than is typically possible in small-scale qualitative studies. By combining the quantitative ratings of activities and achievement grouping with open-ended questions, we also gained insights into students' reasons behind their quantitative evaluations, although small-scale qualitative studies can study this in more depth.

This study did not examine whether the way in which differentiation and achievement grouping were implemented affected students' perceptions. For example, the quality of differentiation - i.e., the degree to which adaptations are carefully matched to students' educational needs - may also affect students' perceptions. In the current study, students' achievement group placement did not always correspond with their achievement on the standardised achievement test administered at the end of the previous schoolyear.

While teachers may have created the achievement groups based on other and more recent achievement information (e.g., curriculum-based tests, daily mathematics work), it might also mean that some students were placed in an achievement group that was not appropriate for their achievement level. This might partially explain why some students perceived the work in their achievement group as too easy or too difficult. In addition, the flexibility of the grouping arrangements might affect students' perceptions (Education Endowment Foundation, 2018). Finally, the use of multigrade classes may have implications for teachers' practices (for example, if teachers create achievement groups within each grade level, they will need to divide their attention over more achievement groups than a teacher teaching a single-grade class) which may in turn affect students' perceptions of differentiation and achievement grouping. These would be interesting issues to explore in future research.

This study focused on differentiation in a specific context, namely primary mathematics education in the Netherlands. Due to substantial differences between countries and content areas in the traditions and practices of differentiation and achievement grouping, these results may not be directly generalisable to other countries or other content areas. In the Netherlands, for example, differentiation practices seem to be somewhat similar for reading, in the sense that teachers might offer more or less difficult reading materials or more or less instruction to different subgroups of students, based on their achievement level (although the way in which instruction and practice are adapted to the needs of low-achieving or high-achieving students may be qualitatively different in reading compared to mathematics, because of the different content area and belonging didactical models). However, for other subjects such as science, teachers seem to use different approaches to differentiation (Slim et al., 2022). Compared to other countries, the achievement grouping practices in the Netherlands may be relatively flexible, which might partially explain the relatively positive evaluations compared to previous studies (e.g., Eder, 1983; Gripton, 2020; Marks, 2013; McGillicuddy & Devine, 2020). Future research could not only study the generalisability of the findings across contexts, but could also use naturally occurring differences in the implementation of differentiation and achievement grouping between various countries and domains to investigate how these differences in implementation might affect students' perceptions.

From the current findings in the context of primary mathematics education in the Netherlands, we conclude that students had largely positive attitudes about differentiation and achievement grouping. Students appreciated it when the amount and difficulty of tasks and instruction were adapted to their current achievement level, and did not like either too easy or too difficult work. While the majority of students across achievement groups wanted to retain the achievement grouping system and reported high liking of their achievement group, students placed in low achievement groups reported to learn less from their group and more often had the desire to be in a different group. Didactical considerations such as wanting to learn more or wanting to be challenged seemed to be more prominent in students' reasoning than socioemotional considerations such as the social status associated with an achievement group.

Our findings have the following implications. Many students displayed positive attitudes to learning: they liked to learn more, did not like to be distracted by other students, and wanted to be challenged. Many students specifically mentioned that they liked or wanted to have enrichment tasks. To retain this positive attitude towards learning, we think that it would be helpful to encourage rather than discourage students who want to try more difficult tasks. This relates to the topic of self-regulation, which has been receiving increasing attention in the differentiation literature: ideally, students should be able to co-decide (in collaboration with the teacher) whether they need additional instruction and which tasks they should do (Van Geel et al., 2018). This might reduce negative experiences with work that is perceived as too hard or too easy. Future research could also examine ways in which the perceived benefits of adapting education to students' achievement level can be retained, while reducing potential socioemotional or didactical disadvantages of placement in a low achievement group. This might include ways to adapt instruction and practice to students' educational needs more flexibly based on students' current understanding of specific mathematical content (perhaps, also using adaptive educational software to assist the teacher in making these

choices), as well as variation of grouping arrangements (for example by using heterogeneous groups in situations where students of diverse achievement levels can learn from each other). In such future research, the perspective of students themselves should not be overlooked.

Keypoints

- Students' voices have not often been heard in the debate on differentiation and achievement grouping
- In this mixed-methods study, primary school students (N = 428) evaluated differentiated activities and within-class achievement grouping
- Didactical and socioemotional aspects, and potential differences between students of diverse achievement levels were considered
- Students had largely positive views on differentiated activities and achievement grouping
- But students also perceived disadvantages of placement in a low achievement group

Acknowledgments

We thank all students and teachers who were involved in this study, as well as the reviewers of this manuscript, for their valuable contributions.

References

Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, *24*(4), 387–426. https://doi.org/10.1007/s40593-014-0023-y

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Blok, H. (2004). Adaptief onderwijs: betekenis en effectiviteit [Adaptive education: meaning and effectivity]. *Pedagogische Studiën*, *81*(1), 5–27.

Boaler, J., Wiliam, D., & Brown, M. (2000). Students' experiences of ability grouping - Disaffection, polarisation and the construction of failure. *British Educational Research Journal*, *26*(5), 631–648. https://doi.org/10.1080/713651583

Campbell, T. (2021). In-class 'ability'-grouping, teacher judgements and children's mathematics self-concept: evidence from primary-aged girls and boys in the UK Millennium Cohort Study. *Cambridge Journal of Education*. https://doi.org/10.1080/0305764X.2021.1877619

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

Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. Harper Perennial.

Deunk, M. I., Smale-Jacobse, A. E., de Boer, H., Doolaard, S., & Bosker, R. J. (2018). Effective

differentiation Practices:A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, *24*, 31–54. https://doi.org/10.1016/j.edurev.2018.02.002

Eder, D. (1983). Ability Grouping and Students' Academic Self-Concepts: A Case Study. *The Elementary School Journal*, *84*, 149–161. https://doi.org/10.2307/1001307

Education Endowment Foundation. (2018). *Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit

Expertgroep doorlopende leerlijnen. (2008). *Over de drempels met rekenen: Consolideren, onderhouden, gebruiken en verdiepen.* Expertgroep doorlopende leerlijnen.

Faber, J. M., & Visscher, A. J. (2016). *De effecten van Snappet: Effecten van een adaptief onderwijsplatform op leerresultaten en motivatie van leerlingen. [The effects of Snappet: Effects of an adaptive educational platform on student achievement and motivation]*. Universiteit Twente.

Francis, B., Connolly, P., Archer, L., Hodgen, J., Mazenod, A., Pepper, D., Sloan, S., Taylor, B., Tereshchenko, A., & Travers, M. C. (2017). Attainment Grouping as self-fulfilling prophesy? A mixed methods exploration of self confidence and set level among Year 7 students. *International Journal of Educational Research*, *86*, 96–108. https://doi.org/10.1016/j.ijer.2017.09.001

Gripton, C. (2020). Children's lived experiences of 'ability' in the Key Stage One classroom: life on the 'tricky table.' *Cambridge Journal of Education*, *50*(5), 559–578. https://doi.org/10.1080/0305764X.2020.1745149

Hallam, S., Ireson, J., & Davies, J. (2004). Primary pupils' experiences of different types of grouping in school. *British Educational Research Journal*, *30*(4), 515–533. https://doi.org/10.1080/0141192042000237211

Hargreaves, E., Buchanan, D., & Quick, L. (2021). "Look at them! They all have friends and not me": the role of peer relationships in schooling from the perspective of primary children designated as "lower-attaining." *Https://Doi.Org/10.1080/00131911.2021.1882942*. https://doi.org/10.1080/00131911.2021.1882942

Hart, S. (1992). Differentiation. Part of the problem or part of the solution? *The Curriculum Journal*, *3*(2), 131–142. https://doi.org/10.1080/0958517920030203

Inspectorate of Education. (2019). *Reken-en wiskundeonderwijs aan potentieel hoogpresterende leerlingen [Mathematics education for potentially high-achieving students]*. Inspectie van het Onderwijs.

Janssen, J., Scheltens, F., & Kraemer, J. M. (2005). *Rekenen-wiskunde groep 3-8: handleidingen [Mathematics test grade 1-6: manuals]*. Cito.

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2005). *Wetenschappelijke verantwoording van de toetsen LOVS rekenen-wiskunde voor groep 3 tot en met 8 [Scientific justification of the mathematics tests for grade 1 through 6]*. Cito.

Jerrim, J. (2021). *The association between within-class grouping and children's achievement in mathematics during Year 2, Year 5 and Year 9. School choices report.* Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Within_class_grouping_report_-_final.pdf

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational*

Psychology Review, *19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kirschner, P. A., & Van Merriënboer, J. G. (2013). Do Learners Really Know Best? Urban Legends in Education. *Educational Psychologist*, *48*, 169–183. https://doi.org/10.1080/00461520.2013.804395

Koerhuis, I. (2010). *Rekenen voor kleuters [Mathematics for Kindergarteners]*. Cito.

Koerhuis, I. ., & Keuning, J. (2011). *Wetenschappelijke verantwoording van de toetsen Rekenen voor kleuters voor groep 1 en 2 [Scientific justification of the tests Mathematics for Kindergarteners]*. Cito.

Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means. R package version 1.5 - 2.1*. https://cran.r-project.org/package=emmeans

Linneberg, M. S., & Korsgaard, S. (2019). Coding qualitative data: a synthesis guiding the novice. *Qualitative Research Journal*, *19*(3), 259–270. https://doi.org/10.1108/QRJ-12-2018-0012/FULL/XML

Marks, R. (2013). "The Blue Table Means You Don't Have a Clue": the persistence of fixed-ability thinking and practices in primary mathematics in English schools. *Forum*, *55*(1), 31. https://doi.org/10.2304/forum.2013.55.1.31

Marsh, H. W. (1984). Self-Concept, Social Comparison, and Ability Grouping: A Reply to Kulik and Kulik. In *Source: American Educational Research Journal* (Vol. 21, Issue 4). Winter.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept [Article]. *Journal of Educational Psychology*, *79*(3), 280–295. https://doi.org/10.1037//0022-0663.79.3.280

McGillicuddy, D., & Devine, D. (2020). 'You feel ashamed that you are not in the higher group'—Children's psychosocial response to ability grouping in primary school. *British Educational Research Journal*, *46*(3), 553–573. https://doi.org/10.1002/berj.3595

Murray, T., & Arroyo, I. (2002). Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems. *LNCS*, *2363*, 749–758.

Park, D., Tsukayama, E., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2016). Young children's motivational frameworks and math achievement: Relation to teacher-reported instructional practices, but not teacher theory of intelligence. *Journal of Educational Psychology*, *108*(3), 300–313. https://doi.org/10.1037/edu0000064

Prast, E.J. & Hickendorff, M. (in press). How do Dutch teachers implement differentiation in primary mathematics education? In: R. Maulana, M. Helms-Lorenz, & R.M. Klassen (Eds.). *Effective teaching around the world: Theoretical, empirical, methodological and practical insights.* Springer.

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*(2), 90–116. https://doi.org/10.14768/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*. https://doi.org/10.1016/j.learninstruc.2018.01.009

Raveaud, M. (2005). Hares, tortoises and the social construction of the pupil: Differentiated learning in French and English primary schools. *British Educational Research Journal*, *31*(4), 459–479. https://doi.org/10.1080/01411920500148697

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*(11), 1186–1204. https://doi.org/10.1080/13603116.2012.743604

Rubie-Davies, C. M. (2014). *Becoming a high expectation teacher: Raising the bar*. Routledge.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.

Slim, T., Van Schaik, J., Hotze, A. & Raijmakers, M. (2022, July). *Differentiatie in het wetenschap & technologie-onderwijs: overtuiging en praktijk van aankomende- en expertleerkrachten. [Differentiation in science & technology education: attitudes and practices of pre-service teachers and expert teachers]*. Poster presented at the Onderwijs Research Dagen [Educational Research Days], Hasselt, Belgium.

Tereshchenko, A., Francis, B., Archer, L., Hodgen, J., Mazenod, A., Taylor, B., Pepper, D., & Travers, M. C. (2019). Learners' attitudes to mixed-attainment grouping: examining the views of students of high, middle and low attainment. *Research Papers in Education*, *34*(4). https://doi.org/10.1080/02671522.2018.1452962

Tieso, C. L. (2003). Ability grouping is not just tracking anymore. *Roeper Review*, *26*(1), 29–36.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., Conover, L. A., & 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. https://doi.org/10.1177/016235320302700203

Van den Bergh, L. (2018). *Waarderen van diversiteit in het onderwijs [Valuing diversity in education]*. Fontys Opleidingscentrum Speciale Onderwijszorg.

Van Geel, M., Keuning, T., Frèrejean, J., Dolmans, D., Van Merriënboer, J., & Visscher, A. J. (2018). Capturing the complexity of differentiated instruction. *School Effectiveness and School Improvement*, *30*(1), 51–67. https://doi.org/10.1080/09243453.2018.1539013

Van Geel, M., Keuning, T., Visscher, A. J., & Fox, J. P. (2016). Assessing the Effects of a School-Wide Data-Based Decision-Making Intervention on Student Achievement Growth in Primary Schools. *American Educational Research Journal*, *53*(2), 360–394. https://doi.org/10.3102/0002831216637346

Vaughn, S., Schumm, J. S., Klingner, J., & Saumell, L. (1995). Students' views of instructional practices: Implications for inclusion. *Learning Disability Quarterly*, *18*(3), 236–248. https://doi.org/10.2307/1511045

Vaughn, S., Schumm, J. S., Niarhos, F. J., & Daugherty, T. (1993). What do students think when teachers make adaptations? *Teaching and Teacher Education*, *9*(1), 107–118. https://doi.org/10.1016/0742-051X(93)90018-C

Vaughn, S., Schumm, J. S., Niarhos, F. J., & Gordon, J. (1993). Students' perceptions of two hypothetical teachers' instructional adaptations for low achievers. *The Elementary School Journal*, *94*(1), 87–102.

Visscher, A. J. (2015). *Over de zin van opbrengstgericht(er) werken in het onderwijs [About the value of (more) data-based decision making in education]*. GION.

Vogt, F., & Rogalla, M. (2009). Developing Adaptive Teaching Competency through coaching. *Teaching and Teacher Education, 25*(8), 1051–1060. https://doi.org/10.1016/j.tate.2009.04.002

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer Verlag New York.

Supplementary materials

Overview of appendices:

Appendix 1: Correspondence between achievement test scores and achievement group placement
Appendix 2: Additional information about the analyses in Part 1
Appendix 3: Additional information about the qualitative analyses in Part 2

Appendix 1: Correspondence between achievement test scores and achievement group placement

Table A1

Correspondence between achievement test scores and achievement group placement

Achievement test score	Low achievement group	Average achievement group	High achievement group
I (highest)	2	14	59
II	3	22	21
III	13	30	4
IV	17	10	9
V (lowest)	17	4	6

Table A1 displays the (imperfect) correspondence between students' achievement group placement and their scores on the achievement test, is based on 231 students who had been placed in a single within-class achievement group during the past three weeks and for whom achievement test scores were available. Note that the achievement test scores were collected at the end of the previous school year. Reasons for non-correspondence may include the assignment to achievement groups based on other measures and more recent sources of information (e.g., curriculum-based tests, students' responses during mathematics lessons).

Appendix 2: Additional information about the analyses in Part 1

Descriptive statistics of the outcome variables are provided in Tables A2 (frequency), A3 (liking), and A4 (learning).

Table A2

Means and standard deviations of student-reported frequency of activities split by achievement level

	Achievement level											
	V (lowest)		IV		III		II		I (highest)		Overall	
Activity	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Whole-class instruction	4.22	1.01	3.96	0.96	4.07	0.99	3.89	0.99	3.83	1.16	3.96	1.05
Working independently	4.02	0.94	3.80	1.21	3.95	1.11	4.23	0.93	4.28	0.86	4.09	1.01
Working together	3.66	1.22	3.18	1.11	3.05	1.06	3.16	1.00	3.20	1.21	3.22	1.14
Easier tasks	3.41	1.41	2.65	1.39	2.70	1.43	2.14	1.44	2.25	1.56	2.54	1.51
Subgroup extended instruction	3.20	1.36	3.05	1.39	2.63	1.40	1.98	1.09	1.86	1.10	2.42	1.36
Individual extended instruction	3.10	1.32	2.64	1.21	2.61	1.35	2.30	1.24	2.04	1.31	2.44	1.33
Enrichment tasks	2.58	1.65	2.64	1.60	2.85	1.62	3.62	1.57	4.09	1.30	3.31	1.64
Subgroup enrichment instruction	2.80	1.56	2.22	1.34	2.15	1.44	2.26	1.32	2.45	1.34	2.36	1.39
Individual enrichment instruction	2.62	1.61	2.18	1.45	2.19	1.41	2.37	1.37	2.42	1.37	2.35	1.42
Overall	3.30	1.46	2.93	1.43	2.91	1.47	2.88	1.47	2.93	1.53		

Note. Students rated the frequency of activities on a five-point scale (5 = highest).

Table A3

Means and standard deviations of student-reported liking of activities split by achievement level

	Achievement level											
	V (lowest)		IV		III		II		I (highest)		Overall	
Activity	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Whole-class instruction	3.80	1.19	3.58	1.27	3.27	1.16	3.30	1.15	3.22	1.27	3.39	1.23
Working independently	3.44	1.43	3.45	1.58	3.53	1.33	3.89	1.32	3.78	1.27	3.65	1.38
Working together	4.20	1.17	4.05	1.37	4.30	1.01	3.98	1.17	3.89	1.37	4.05	1.25
Easier tasks	3.80	1.38	3.25	1.57	3.10	1.54	2.46	1.66	2.38	1.58	2.88	1.63
Subgroup extended instruction	3.88	1.36	3.27	1.46	3.29	1.50	2.61	1.50	2.35	1.27	2.95	1.50
Individual extended instruction	3.50	1.55	2.84	1.51	2.58	1.42	2.45	1.33	2.34	1.30	2.65	1.44
Enrichment tasks	2.52	1.57	3.20	1.59	3.10	1.64	4.11	1.30	4.15	1.25	3.56	1.56
Subgroup enrichment instruction	3.15	1.67	3.02	1.50	2.83	1.53	3.16	1.33	3.05	1.35	3.04	1.45
Individual enrichment instruction	2.88	1.62	2.89	1.52	2.41	1.43	2.88	1.36	2.74	1.40	2.75	1.45
Overall	3.47	1.52	3.29	1.52	3.16	1.49	3.20	1.48	3.10	1.49		

Note. Students rated their liking of activities on a five-point scale (5 = highest).

Table A4

Means and standard deviations of student-reported learning from activities split by achievement level

	Achievement level											
	V (lowest)		IV		III		II		I (highest)		Overall	
Activity	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Whole-class instruction	4.15	1.06	3.93	1.26	3.48	1.30	3.75	1.18	3.70	1.11	3.77	1.19
Working independently	3.85	1.20	3.84	1.23	4.18	0.91	4.23	0.87	4.10	0.88	4.06	1.00
Working together	4.10	1.09	3.80	1.21	3.92	1.06	3.74	1.19	3.82	1.10	3.86	1.13
Easier tasks	3.27	1.45	2.89	1.50	2.69	1.51	2.09	1.34	2.20	1.43	2.54	1.50
Subgroup extended instruction	4.17	1.18	4.09	1.14	3.72	1.20	3.33	1.50	3.43	1.37	3.69	1.33
Individual extended instruction	3.95	1.20	4.00	1.22	3.71	1.30	3.51	1.35	3.57	1.37	3.71	1.31
Enrichment tasks	3.33	1.61	4.24	1.04	4.12	1.14	4.52	1.04	4.57	0.66	4.25	1.12
Subgroup enrichment instruction	3.74	1.52	3.69	1.36	3.73	1.28	3.72	1.24	3.82	1.28	3.75	1.31
Individual enrichment instruction	3.65	1.53	3.71	1.33	3.49	1.32	3.68	1.31	3.65	1.31	3.64	1.34
Overall	3.80	1.35	3.80	1.30	3.67	1.29	3.62	1.38	3.65	1.33		

Note. Students rated their learning from activities on a five-point scale (5 = highest).

Analytical strategy

Data were analysed in R with multilevel models to take into account the nested data structure (e.g., scores of students within a class/school being correlated). For all analyses, in order to take the dependencies in the data due to the nested data structure into account, we used random intercepts at the various levels (i.e., student, class and school). As a four-level random intercept model is already quite complex for our data, we decided not to include random slopes in the model. The models were fitted with the package LME4 (Bates et al., 2015), post-hoc analyses were conducted with the package EMMEANS (Lenth, 2020), and figures were plotted with GGPLOT2 (Wickham, 2016). Effect coding was used. Effect coding differs from dummy

coding in the sense that weights other than 0 and 1 (i.e., standard dummy coding) can be assigned to the various levels of a categorical variable, which facilitates the interpretation of the fixed effects.

First, we estimated an empty model – also called unconditional means model (Singer & Willett, 2003) - to investigate the amount of variance at the various levels. Second, main effects of activity and achievement level were added to the model. Third, the interaction between activity and achievement level was added. To evaluate the significance of main effects and interaction effects, Likelihood Ratio Tests (LRT) were used to compare the fit of the full model (i.e., including the effect of interest) to the fit of a reduced model without that main effect or interaction.

Empty models

The empty models indicated that by far the most variance was at the level of the various activities rated by the same students (i.e., repeated measures, see Table A5). The variance at the student level was somewhat larger for the degree to which students perceived to learn from activities (13.1%) than for the other outcome variables. The amount of variance at the class and school level was quite small (0.5 – 2.2%), but these levels were retained in the analyses anyway to correct for any clustering effects at these levels.

Table A5

Distribution of the outcome variance across the different levels in the data

Level	**Frequency**	**Liking**	**Learning**
Activity (within-student)	92.7%	90.6%	84.8%
Student	4.0%	5.8%	13.1%
Class	1.4%	2.2%	1.6%
School	1.9%	1.4%	0.5%

Likelihood ratio tests comparing model fit

Table A6 provides an overview of the results of the likelihood ratio tests comparing model fit. As can be seen in the table, the model including the interaction between activity and achievement level had the best fit compared to a reduced model for all outcome variables.

Table A6

Outcomes of Likelihood ratio tests comparing model fit

	Frequency			Liking			Learning		
Model comparison	$\chi 2$	*df*	*p*	$\chi 2$	*df*	*p*	$\chi 2$	*df*	*p*
Main effect of activity added to empty model	621.79	8	<.001	248.03	8	<.001	326.19	8	<.001
Main effect of achievement added to previous model	4.55	1	0.03	6.36	1	0.01	2.05	1	0.152
Interaction effect added to previous model	182.36	8	<.001	161.75	8	<.001	95.74	8	<.001

Post-hoc tests for the interaction effects

Table A7 provides an overview of the post-hoc tests for the interaction effect. A significant effect means that students' achievement level predicts their ratings for that activity. Since a score of 1 on the achievement tests reflects the highest achievement and 5 the lowest, a positive value for the effect indicates a negative effect of achievement level (i.e., activity ratings are higher for low-achieving students) whereas a negative value indicates a positive effect of achievement level (i.e., activity ratings are higher for high-achieving students).

Table A7

Post-hoc tests of the interaction effect: the effect of achievement on students' reported frequency, liking and learning for each activity separately

Activity and outcome variable	$\hat{\beta}$	*SE*	*df*	*t*	*p*
Whole-class instruction					
Frequency	-0.07	0.05	2491	-1.50	0.13
Liking	-0.12	0.06	2514	-2.22	0.03
Learning	-0.08	0.05	2126	-1.62	0.11
Working independently					
Frequency	0.11	0.05	2488	2.16	0.03
Liking	0.12	0.06	2511	2.15	0.03
Learning	0.08	0.05	2126	1.64	0.10
Working together					
Frequency	-0.06	0.05	2491	-1.23	0.22
Liking	-0.07	0.06	2511	-1.22	0.22
Learning	-0.04	0.05	2131	-0.86	0.39
Easier tasks					
Frequency	-0.25	0.05	2488	-4.88	<.001
Liking	-0.34	0.06	2512	6.11	<.001
Learning	-0.27	0.05	2142	5.08	<.001
Subgroup extended instruction					
Frequency	-0.37	0.05	2488	-7.23	<.001
Liking	-0.36	0.06	2515	-6.43	<.001
Learning	-0.11	0.05	2142	5.08	<.001
Individual extended instruction					
Frequency	-0.23	0.05	2496	-4.63	<.001
Liking	-0.23	0.06	2522	-4.16	<.001
Learning	-0.11	0.05	2148	-2.31	<.001
Enrichment tasks					
Frequency	0.43	0.05	2496	8.40	<.001
Liking	0.41	0.06	2521	7.37	<.001
Learning	0.25	0.05	2142	5.08	<.001
Subgroup enrichment instruction					
Frequency	-0.02	0.05	2495	-0.44	0.66
Liking	0.01	0.06	2517	0.27	0.79
Learning	0.03	0.05	2156	0.61	0.54
Individual enrichment instruction					
Frequency	0.01	0.05	2494	0.15	0.88
Liking	-0.01	0.06	2520	-0.15	0.88
Learning	0.01	0.05	2137	0.15	0.88

Appendix 3: Additional information about the qualitative analyses in Part 2

Tables A8 through A11 provide an overview of the answering categories (lower-order codes organised by higher-order themes) for each question, as well as the number of times these categories were mentioned by students placed in low, average and high achievement groups.

Table A8

Students' responses to the question: What do you like about being in your achievement group?

Answering category	Example	Frequency of comments [a] Low (*n* = 59)	Average (*n* = 97)	High (*n* = 110)	Total (*N* = 266)
Answers about work / difficulty		***17 (28.8%)***	***32 (33.0%)***	***54 (49.1%)***	***103 (38.7%)***
Appropriate difficulty	You don't need to do too difficult or too easy stuff	8 (13.6%)	15 (15.5%)	29 (26.4%)	52 (19.6%)
Challenge	Because if I have this work, it's a challenge	1 (1.7%)	2 (2.1%)	13 (11.8%)	16 (6.0%)
Working more / doing many tasks	So you can start to work immediately	3 (5.1%)	6 (6.2%)	4 (3.6%)	13 (4.9%)
Tasks / activities are fun	You get nice tasks	1 (1.7%)	1 (1.0%)	7 (6.4%)	9 (3.4%)
Fewer tasks and/or finish earlier	I need to do less and then I finish earlier	4 (6.8%)	3 (3.1%)	1 (0.9%)	8 (3.0%)
My level	It matches my level	0 (0.0%)	5 (5.2%)	0 (0.0%)	5 (1.9%)
Answers about group		***17 (28.8%)***	***15 (15.5%)***	***23 (20.9%)***	***55 (20.7%)***
Positive about group members / dynamics	The children in this group are kind	8 (13.6%)	7 (7.2%)	13 (11.8%)	28 (10.5%)
Helping each other / working together	Because if you don't know something the other children can help you	5 (8.5%)	4 (4.1%)	5 (4.6%)	14 (5.3%)
Concentrate / no distraction	Because it's a quiet group and I can concentrate better	4 (6.8%)	2 (2.1%)	2 (1.8%)	8 (3.0%)
Group rank: being in a high(er) / smart(er) group	Then I think sometimes that I am a bit smarter together with the other children and that gives me confidence	0 (0.0%)	2 (2.1%)	3 (2.7%)	5 (1.9%)
Answers about learning		***7 (11.9%)***	***12 (12.4%)***	***11 (10.0%)***	***30 (11.3%)***
Learn more / faster / new things	Then you learn more of it	4 (6.8%)	8 (8.3%)	9 (8.2%)	21 (7.9%)
Understanding better	Then you understand better	2 (3.4%)	3 (3.1%)	0 (0.0%)	5 (1.9%)
Getting smarter / better at math / higher level	Then you get super smart	1 (1.5%)	1 (1.0%)	2 (1.8%)	4 (1.5%)
Answers about instruction / teacher		***5 (8.5%)***	***11 (11.3%)***	***8 (7.3%)***	***24 (9.0%)***
(More) explanation /	That you get more help	4	7	3	14

instruction / help		(6.8%)	(7.2%)	(2.7%)	(5.3%)
Positive about teacher	The teacher thinks of a nice way	0 (0.0%)	2 (2.1%)	4 (3.6%)	6 (2.3%)
No (additional) instruction / explanation	You don't need to listen to another explanation	1 (1.7%)	2 (2.1%)	1 (0.9%)	4 (1.5%)
General and other answers		***13 (22.0%)***	***27 (27.8%)***	***14 (12.7%)***	***54 (20.3%)***
Unspecific / don't know / other	I just like it	10 (17.0%)	17 (17.5%)	9 (8.2%)	36 (13.5%)
Nothing is nice / negative comment	I don't like it at all	3 (5.1%)	7 (7.2%)	3 (2.7%)	13 (4.9%)
Everything is nice	Everything about this group is nice	0 (0.0%)	3 (3.1%)	2 (1.8%)	5 (1.9%)

[a] Number and percentage of comments within that achievement group. The total number of comments exceeds the number of students since some answers belonged to two categories.

Table A9

Students' responses to the question: What don't you like about being in your achievement group?

Answering category	**Example**	**Frequency of comments[a] Low (*n* = 55)**	**Average (*n* = 89)**	**High (*n* = 105)**	**Total (*N* = 249)**
Answers about work / difficulty		***9 (16.4%)***	***15 (16.9%)***	***19 (18.1%)***	***43 (17.3%)***
Needing to work much / longer / fast	You need to finish quickly	1 (1.8%)	3 (3.4%)	11 (10.5%)	15 (6.0%)
Inappropriate difficulty: too hard	Sometimes it's difficult	1 (1.8%)	5 (5.6%)	6 (5.7%)	12 (4.8%)
Boring / takes a long time	Because it takes a super long time. I find it super boring.	2 (3.6%)	4 (4.5%)	1 (1.0%)	7 (2.8%)
Inappropriate difficulty: too easy	Because it's too easy now	4 (7.3%)	1 (1.1%)	1 (1.0%)	6 (2.4%)
Want more challenge	Because I like to be challenged	1 (1.8%)	2 (2.3%)	0 (0.0%)	3 (1.2%)
Answers about group		***9 (16.4%)***	***8 (9.0%)***	***20 (19.1%)***	***37 (14.9%)***
Distraction	I am distracted when the others talk	4 (7.3%)	3 (3.4%)	12 (11.4%)	19 (7.6%)
Negative about group members / dynamics	We quarrel sometimes	2 (3.6%)	2 (2.3%)	4 (3.8%)	8 (3.2%)
Group rank: Want to be in higher group / not nice to be in low group	Of course, I would rather be in the plus-group [= highest group]	3 (5.5%)	3 (3.4%)	0 (0.0%)	6 (2.4%)
Stress about high achievement group	The stress	0 (0.0%)	0 (0.0%)	4 (3.8%)	4 (1.6%)
Answers about learning		***3 (5.4%)***	***5 (5.6%)***	***1 (1.0%)***	***9 (3.6%)***
Not understanding	When I don't understand	1 (1.8%)	3 (3.4%)	1 (1.0%)	5 (2.0%)

Learning less / not much	I don't learn so much and I want to get better	1 (1.8%)	1 (1.1%)	0 (0.0%)	2 (0.8%)
Being (called) bad / stupid	It's not nice to be so bad	1 (1.8%)	1 (1.1%)	0 (0.0%)	2 (0.8%)
Answers about instruction / teacher		***2 (3.6%)***	***2 (2.3%)***	***4 (3.8%)***	***8 (3.2%)***
Negative about instruction / teacher	You get additional explanation when you understand it already	2 (3.6%)	2 (2.3%)	4 (3.8%)	8 (3.2%)
General and other answers		***32 (58.2%)***	***59 (66.3%)***	***61 (58.1%)***	***152 (61.0%)***
Everything is nice / positive comment	There's nothing about this group that I don't like	12 (21.8%)	34 (38.2%)	41 (39.1%)	87 (35.0%)
Unspecific / don't know / other	I just don't like it	20 (36.4%)	25 (28.0%)	20 (19.0%)	65 (26.1%)

[a] Number and percentage of comments within that achievement group. The total number of comments exceeds the number of students since some answers belonged to two categories.

Table A10

Students' responses to the question: Why do you learn much or little of being in your achievement group?

		Frequency of comments [a]			
Answering category	**Example**	**Low (*n* = 52)**	**Average (*n* = 90)**	**High (*n* = 104)**	**Total (*N* = 246)**
Answers about work / difficulty		***9 (17.3%)***	***17 (18.9%)***	***42 (40.4%)***	***68 (27.6%)***
Difficult (positive) / challenging	You learn much because you also get more difficult sums	2 (3.9%)	4 (4.4%)	30 (28.9%)	36 (14.6%)
Working independently	I can do it by myself	1 (1.9%)	4 (4.4%)	3 (2.9%)	8 (3.3%)
Appropriate for my level	It's my level	0 (0.0%)	4 (4.4%)	3 (2.9%)	7 (2.9%)
Medium or varying difficulty	Sometimes it's easy and sometimes it's difficult	2 (3.9%)	2 (2.2%)	2 (1.9%)	6 (2.4%)
Too easy / not challenging / want more challenge	Now it is sometimes a bit too easy	1 (1.9%)	3 (3.3%)	1 (1.0%)	5 (2.0%)
Too difficult	…and I don't learn much because it's sometimes difficult	1 (1.9%)	0 (0.0%)	2 (1.9%)	3 (1.2%)
Easy/not too difficult	Because the plus-group [highest group] would be too difficult	2 (3.9%)	0 (0.0%)	1 (1.0%)	3 (1.2%)
Answers about learning		***12 (23.1%)***	***13 (14.4%)***	***12 (11.5%)***	***37 (15.0%)***
Learn more / specific math content / getting more sums	You learn new goals every time	3 (5.8%)	7 (7.8%)	6 (5.8%)	16 (6.5%)
Understand more / better	I learn much because I understand it better now	3 (5.8%)	5 (5.6%)	3 (2.9%)	11 (4.5%)

Get smarter / better at math	Because I get smarter	1 (1.9%)	1 (1.1%)	2 (1.9%)	4 (1.6%)
Not understanding	Because I don't understand it at all	3 (5.8%)	0 (0.0%)	1 (1.0%)	4 (1.6%)
Learn less / fewer tasks	Because you do get fewer sums	2 (3.9%)	0 (0.0%)	0 (0.0%)	2 (0.8%)
Answers about instruction / teacher		***7 (13.5%)***	***17 (18.9%)***	***12 (11.5%)***	***36 (14.6%)***
Positive about instruction / teacher	Because the teacher gives you more explanations and does more sums with you	7 (13.5%)	15 (16.7%)	12 (11.5%)	34 (13.8%)
Negative about instruction / teacher	I learn more from myself because the teachers confuse me	0 (0.0%)	2 (2.2%)	0 (0.0%)	2 (0.8%)
Answers about group		***5 (9.6%)***	***7 (7.8%)***	***7 (6.7%)***	***19 (7.7%)***
Helping each other / working together	When you work together you also learn from it	3 (5.8%)	1 (1.1%)	3 (2.9%)	7 (2.9%)
Distraction	I learn a bit less because sometimes children talk and draw attention	2 (3.9%)	1 (1.1%)	3 (2.9%)	6 (2.4%)
Positive about group members / dynamics	Because it's nice in a nice group	0 (0.0%)	3 (3.3%)	0 (0.0%)	3 (1.2%)
Concentrate / No distraction	Quiet so more concentration	0 (0.0%)	2 (2.2%)	1 (1.0%)	3 (1.2%)
General and other answers		***19 (36.5%)***	***35 (38.9%)***	***31 (29.8%)***	***86 (35.0%)***
Unspecific / don't know / other	I learn much because I learn much from it	17 (32.7%)	34 (37.8%)	29 (27.9%)	80 (32.5%)
Fun	Because I think it's fun	2 (3.9%)	1 (1.1%)	2 (1.9%)	6 (2.4%)

[a] Number and percentage of comments within that achievement group. The total number of comments exceeds the number of students since some answers belonged to two categories.

Table A11

Students' responses to the question: Why would you (not) prefer to be in a different achievement group?

		Frequency of comments[a]			
Answering category	**Example**	**Low (*n* = 54)**	**Average (*n* = 87)**	**High (*n* = 106)**	**Total (*n* =247)**
Reasons for preferring to stay in the same group					
Answers about work / difficulty		***6 (11.1%)***	***15 (17.2%)***	***36 (34.0%)***	***57 (23.1%)***
Difficulty [appropriate in current group]	Because I don't do too difficult or too easy work	5 (9.3%)	9 (10.3%)	22 (20.8%)	36 (14.6%)
Tasks or activities in the current group are fun / want to keep	Enrichment tasks are fun, so I want to keep those	1 (1.9%)	3 (3.5%)	5 (4.7%)	9 (3.6%)

enrichment tasks					
My level	Because this is my level	0 (0.0%)	3 (3.5%)	3 (2.8%)	6 (2.4%)
Challenge	Because otherwise I don't have enough challenge	0 (0.0%)	0 (0.0%)	6 (5.7%)	6 (2.4%)
Answers about group		***6 (11.1%)***	***3 (3.5%)***	***6 (5.7%)***	***15 (6.1%)***
Group members / dynamics / no distraction	They are kind and not loud	5 (9.3%)	3 (3.5%)	5 (4.7%)	12 (4.9%)
Group rank: being in a high(er) / smart(er) group	It's the best group	1 (1.9%)	0 (0.0%)	1 (0.9%)	2 (0.8%)
Answers about learning		***1 (1.9%)***	***4 (4.6%)***	***5 (4.7%)***	***10 (4.1%)***
Learn more / faster / new things	Because I learn most in this group	1 (1.9%)	4 (4.6%)	3 (2.8%)	8 (3.2%)
Getting smarter / better at math	Because I get smarter	0 (0.0%)	0 (0.0%)	2 (1.9%)	2 (0.8%)
Answers about instruction / teacher		***0 (0.0%)***	***2 (2.3%)***	***4 (3.8%)***	***6 (2.4%)***
Positive about instruction / teacher	Because now I get the explanation that I need	0 (0.0%)	2 (2.3%)	4 (3.8%)	6 (2.4%)
General and other answers		***14 (25.0%)***	***34 (39.1%)***	***44 (41.5%)***	***92 (37.2%)***
This group is nice/fun	Because it's nice in this group	6 (11.1%)	23 (26.4%)	20 (18.9%)	49 (19.8%)
Unspecific / don't know / other	I find it hard to explain	8 (14.8%)	11 (12.6%)	24 (22.6%)	43 (17.4%)
<u>Reasons for preferring to be in a different group</u>					
Answers about work / difficulty		***6 (11.1%)***	***8 (9.2%)***	***4 (3.8%)***	***18 (7.3%)***
Difficulty [would be more appropriate in another group]	More difficult sums: because I like that	3 (5.6%)	4 (4.6%)	4 (3.8%)	11 (4.5%)
Tasks or activities in the other group are fun / want to get enrichment tasks	Because I want enrichment tasks	2 (3.7%)	2 (2.3%)	0 (0.0%)	4 (1.6%)
Challenge	More challenge	1 (1.9%)	2 (2.3%)	0 (0.0%)	3 (1.2%)
Answers about group		***7 (13.0%)***	***5 (5.8%)***	***2 (1.9%)***	***14 (5.7%)***
Group members / dynamics /distraction	Because I have many friends in the other group	5 (9.3%)	2 (2.3%)	1 (0.9%)	8 (3.2%)
Group rank: being in a high(er) /	Because I want to be in a higher group	2 (3.7%)	3 (3.5%)	1 (0.9%)	6 (2.4%)

smart(er) group					
Answers about learning		***5 (9.3%)***	***2 (2.3%)***	***0 (0.0%)***	***7 (2.8%)***
Learn more / faster / new things	Because I learn much	1 (1.9%)	0 (0.0%)	0 (0.0%)	1 (0.4%)
Getting smarter / better at math / higher level	Then I think I will get better at math	4 (7.4%)	2 (2.3%)	0 (0.0%)	6 (2.4%)
Answers about instruction / teacher		***1 (1.9%)***	***1 (1.2%)***	***0 (0.0%)***	***2 (0.8%)***
Instruction / teacher	Then the teacher can help me when I find it difficult	1 (1.9%)	1 (1.2%)	0 (0.0%)	2 (0.8%)
General and other answers		***8 (14.8%)***	***13 (14.9%)***	***5 (4.7%)***	***26 (10.5%)***
Other group is nice/fun or current group is not nice	It's not as much fun as another group	4 (7.4%)	2 (2.3%)	1 (0.9%)	7 (2.8%)
Unspecific / don't know / other	I can't really explain	4 (7.4%)	11 (12.6%)	4 (3.8%)	19 (7.7%)

[a] Number and percentage of comments within that achievement group. The total number of comments exceeds the number of students since some answers belonged to two categories.

Table A12

Students' responses to the question: Why would you prefer to work with or without achievement groups?

		Frequency of comments[a]			
Answering category	**Example**	**Low (*n* = 52)**	**Average (*n* = 87)**	**High (*n* = 102)**	**Total (*n* = 241)**
Reasons for preferring to retain groups		***37 (71.2%)***	***67 (77.0%)***	***80 (78.4%)***	***184 (76.4%)***
Positive about groups / like it as it is	Because I like to work in a group	13 (25.0%)	29 (33.3%)	18 (17.7%)	60 (24.9%)
Between-student differences / appropriate difficulty	Because if you all work at the same level some people don't get explanation and others don't get challenge	6 (11.5%)	14 (16.1%)	24 (23.5%)	44 (18.3%)
Learning / working better with groups	Because you can learn better like this	2 (3.9%)	8 (9.2%)	13 (12.8%)	23 (9.5%)
Possibility to get more instruction	I like it when you can choose whether you want explanation	4 (7.7%)	2 (2.3%)	3 (2.9%)	9 (3.7%)
Know your level	It's much more fun when everybody knows in which star [=group] they are	1 (1.9%)	1 (1.5%)	1 (1.0%)	3 (1.2%)
Unspecific / don't know / other	I don't know how without [groups]	11 (21.2%)	13 (14.9%)	21 (20.6%)	45 (18.7%)
Reasons for preferring to work without groups		***15 (28.9%)***	***20 (23.0%)***	***22 (21.6%)***	***57 (23.7%)***
Learning / working better	Then you can do all tasks and then you get smarter	4 (7.7%)	7 (8.0%)	4 (3.9%)	15 (6.2%)

without groups					
Negative about groups	It's so complicated now with all those groups	2 (3.9%)	2 (2.3%)	4 (3.9%)	8 (3.3%)
Equality, everybody should be/do the same	Because I think that everybody should get the same tasks	2 (3.9%)	3 (3.5%)	1 (1.0%)	6 (2.5%)
Unspecific / don't know / other	Because I want that	7 (13.5%)	8 (9.2%)	13 (12.8%)	28 (11.6%)

[a] Number and percentage of comments within that achievement group. The total number of comments exceeds the number of students since some answers belonged to two categories.