The Effect of Group Formation and Differentiated Instruction on Learning Outcomes. Evidence from Two Randomised Experiments.[1]

Kaat Iterbeke, Kristof De Witte, Wouter Schelfhout, and Koen Declercq

Abstract. We study how group formation and differentiated instruction affect learning outcomes. In particular, the effect of within-class heterogeneous and homogeneous group formations and differentiated instruction are evaluated on the proficiency level of eighth and ninth grade students in a financial education programme. Our paper provides evidence on both effects using two experiments involving 69 schools and 2,407 students. The identification strategy relies on a random allocation of schools to experimental conditions. The results suggest that, overall, the programme increases student financial proficiency by almost a quarter of a standard deviation. Although no main effects of group formations and differentiated instruction are found, the effect is heterogeneous as low ability students significantly benefit from homogeneous group formation and differentiated instruction.

Keywords: Economics of Education, Financial Literacy, Group Formation, Differentiated Instruction, Randomised Experiment

JEL-classification: C93, I21, O16

[1] We acknowledge financial support from the Flemish Science Organisation.

1. Introduction

Today's classrooms have become increasingly diverse. Given that students with diverse learning profiles, cultural and economic backgrounds are seated together in classrooms, teachers are forced to revise their way of teaching. While changing the degree of heterogeneity within classrooms and implementing differentiated instruction could resolve for the issue, classroom diversity, in turn, creates the possibility of positive peer- and knowledge spillover effects because of potential complementarities in abilities (Betts & Shkolnik, 2000; Duflo, Dupas, & Kremer, 2011; Figlio & Page, 2002). This paper examines the consequences of similar education policies by exploring the effect of two differentiation practices, i.e. within-class group formations and differentiated instruction, on student learning outcomes in a randomised evaluation. Examining these effects is important from a policy perspective as manipulating the class composition and adopting differentiated instruction in order to cope with the prevailing heterogeneity may increase the effectiveness of the education production function at distinct lower costs than other policies such as class size reductions (Giorgi, Pellizzari, & Woolston, 2012).

Research suggests classroom diversity to be pronounced in financial literacy education as well (Lusardi & Mitchell, 2014). That is, due to the prevailing heterogeneity in students' opportunities available to improve financial literacy, which are, among others, linked to gender, socio-economic status and immigrant background, performances in financial literacy differ widely (OECD, 2017).

Financial literacy has been considered as a skill essential for participation in today's society due to the increasingly complex and challenging financial landscape. It is defined as a combination of the knowledge of financial concepts and risks on the one hand, and the skills and motivation on the other hand, to apply such knowledge in order to make effective financial decisions, and hence, improve the financial well-being of individuals and society (OECD, 2017). High levels of financial literacy are, among others, positively correlated with wealth accumulation, retirement planning and debt management (van Rooij, Lusardi, & Alessie, 2011). Nevertheless, despite its importance, the literature shows pervasive low levels of financial literacy around the world (Lusardi & Mitchell, 2014). Accordingly, governments, non-profit and financial organisations have been promoting the development of financial education programmes. Although former remedial approaches targeted adults, the focus has shifted towards school-based interventions. Apart from financial illiteracy being more pronounced among youth, school-based programmes are attractive for a number of reasons. Not only can financial guidance be ensured at the appropriate time, i.e. students are still developing habits and are used to absorb, recall and apply learning on regular basis, school-based programmes also offer the possibility to minimise participation issues and generate spillover effects as students act as agents of change in their household's financial decisions (Frisancho, 2018; Lusardi, Mitchell, & Curto, 2010; Mandell, 2008).

The widespread financial illiteracy calls for causal inference on the effectiveness of financial education initiatives (OECD, 2017). Accordingly, a growing literature on the impact of financial education among youth can be found (such as Becchetti, Caiazza, & Coviello, 2013; Becchetti & Pisani, 2012; Berry, Karlan, & Pradhan, 2018; Bruhn, de Souza Leão, Legovini, Marchetti, & Zia, 2016; Frisancho, 2018; Hinojosa et al., 2010; Lührmann, Serra-Garcia, & Winter, 2015; Villanueva, Bover, & Hospido, 2018). Among randomised evaluations, one important contribution is made by Bruhn et al. (2016). With a representative sample of about 25,000 students, the authors examined the impact of a school-based financial education programme with 15-17 year-old high school students in Brazil. They found positive treatment effects on students' performance in a financial proficiency test, yet mixed effects on short-term financial behaviours, such that the importance of distinguishing effects by the dimensions of financial knowledge and behaviour is confirmed.

It is important to note that previous evaluations of school-based financial literacy interventions do not consider heterogeneity of students' performances in financial literacy at the planning phase of the programmes, i.e. all students are taught the same programme in a uniform way. Given the increasing heterogeneity in classrooms, a differentiated approach, aimed at improving financial literacy of *all* students, is likely to be more effective.

Differentiating, in contrast to a one-size-fits-all approach, ensures all students in the classroom to have equity of access to high-quality learning (Gamoran & Weinstein, 1998). According to Tomlinson (2001), students may differ in at least three areas in a differentiated classroom, i.e. student readiness, interest and learning profile. To respond to these differences, a set hallmarks for effective differentiation are proposed, among which the use of small flexible groups in the classroom (Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., Conover, L., Reynolds, T, 2003). Although small-group settings are found to give teachers the flexibility to address student variance more appropriately than whole-class instruction, meta-analyses comparing homogeneous and heterogeneous group formations, where methods of instructions were held constant, reveal little or no significant differences (Lou et al., 1996; Slavin, 1990). [2] Yet, different patterns for low, medium and high ability students within classrooms are observed. Apart from the mixed results bringing the implementation of different group formations into debate, they also indicate the presence of other factors playing a role in the effectiveness (Deunk, Doolaard, Smale-Jacobse, & Bosker, 2015). Hence, a second hallmark by Tomlinson et al. (2003), and suggested by others, is the crucial role of adapting instruction and material to the needs of all groups, i.e. the effectiveness of within-class homogeneous group formation is found to increase when combined with differentiated instruction (Lou et al., 1996). That is, teachers should match the materials to the explicit instructional needs of students, such as giving more support to low ability students and more difficult material to high achievers (Slavin, 1990). Note that, despite the substantial body of educational research on differentiation, only little attention has been

[2] Homogeneous group formation refers to students with similar ability working together in smaller groups, whereas heterogeneous group formation refers to students with different abilities engaging in cooperative learning.

drawn to causal inference. Most of the recent research is quasi-experimental, small-scale, based on specific learning focus, groups and certain subject areas.[3] Within the field of financial literacy, the effectiveness of differentiation has even not been examined to date.

This paper contributes to the literature on economics of education in general, and financial literacy education in particular. We set up a large scale randomised experiment to provide causal inference on the effectiveness of a school-based financial education programme including different differentiation practices. These practices consist of homogeneous and heterogeneous within-class group formations and differentiated instruction, for which the material is adapted according to students' ability. Accordingly, new insights with respect to the importance of the learning context and financial literacy delivery methods are provided. In particular, we examine the following research question: *Do within-class group formations and differentiated instruction increase the learning outcomes of students? In particular, what is the importance of within-class group formations and differentiated instruction on the effectiveness of financial literacy tools?* We refer to the two experiments as two separate waves which are in the eighth and ninth grade of secondary school in the northern region of Belgium, Flanders. Although students in Flanders perform above the PISA average in financial literacy, distinct heterogeneity in students' performance is observed within the region. Further, examining the different age groups not only allows us to observe age-specific heterogeneity in the treatment, it can also be used to assess the external validity of results, i.e. if results appear to be similar in both waves, the likelihood increases that our findings can be generalized to other contexts. Finally, using classroom observations, the programme implementation is examined within schools, such that the findings may give us additional insights into the mechanisms.

The remainder of the article unfolds as follows. In the next section, we present the design and the data. Section 3 provides the empirical model. Section 4 and 5 present the results and discuss some potential mechanisms, respectively. Section 6 concludes.

[3] See for example Alavania & Farhady, 2012; Baumgartner, Lipowski, & Rush, 2003; Chamberlin & Powers, 2010; Houtveen, van de Grift, & Creemers, 2004; Simpkins, Mastropieri, & Scruggs, 2009; Tieso, 2005; Valiandes, 2015.

2. Experimental design

We conducted two randomised experiments involving secondary school students aged 13 to 16 years, living in the region of Flanders, Belgium. The final sample includes 2,407 students from 200 classes in 69 schools. The first experimental wave took place between January, 2018 and March, 2018. The sample for this first wave consists of 1,896 students in 148 eighth grade classes of 43 schools. A second experimental wave was implemented in ninth grade and covers data on 511 students in 52 classes of 26 schools. The second wave started at the end of February, 2018 and ended at the beginning of June, 2018.

2.1. Procedure

In an open call, secondary schools were invited to use novel didactic material on financial literacy. From those which signed up, we assigned schools randomly to a control condition and multiple experimental conditions. Randomising at school level ensures contamination and spill-over effects for schools with multiple classes to be minimal. Moreover, to guarantee a similar implementation of the didactic material across all conditions, we requested the teachers and school principals to follow a strict set of terms. Participation in the experiment was conditional on accepting these terms. First, to avoid contamination from parents, the material had to be delivered during regular class hours. Second, to measure the added-value of the intervention, all students had to take three financial literacy tests, i.e. a pre-treatment test prior to the financial education programme, measuring the baseline financial proficiency of students, and two post-treatment tests, capturing potential short- and long-term impacts of the programme. Third, to avoid timing effects, schools had to deliver the material in pre-specified periods. Table I presents the study timeline for treatment and control schools of wave I and II.

Table I: Study timeline for wave I and II

WAVE I	**Oct-Jan'18**	**Jan'18**	**Mid-End Jan'18**	**End Jan'18**		**End Mar'18**
Treatment schools	Registration	Randomization to an experimental condition	Student pre-test (*2 weeks to complete)*	Material sent to schools	*8 weeks to teach 4 hour course*	Student post-test & teacher survey
Control schools	Registration	Randomization to control condition	Student pre-test (*2 weeks to complete)*	/	/	Student post-test & teacher survey
WAVE II	**Jan-Feb'18**	**End Feb'18**	**End Feb-Mar'18**	**Mid Apr'18**		**Jun'18**
Treatment schools	Registration	Randomization to an experimental condition	Student pre-test (*2 weeks to complete)*	Material sent to schools	*6 weeks to teach 4 hour course*	Student post-test & teacher survey
Control schools	Registration	Randomization to control condition	Student pre-test (*2 weeks to complete)*	/	/	Student post-test & teacher survey

Note: Teachers from treatment schools of wave II were given 6 weeks instead of 8 to plan the course due to time restrictions.

The Financial Literacy Programme. To examine the impact of within-class group formations and differentiated instruction in financial literacy education, we designed four conditions: one control condition where schools did not receive the course material, and three experimental conditions. All

experimental conditions made use of very similar didactic materials, which was designed by secondary school teachers as four lectures of 50 minutes on the topic of means of payment in the format of a serious game. We used this format in order to create an interactive environment which was directed to a clear goal and enabled to provide immediate feedback to students. Apart from these practical grounds, research has also indicated serious games to be more effective in terms of learning and retention compared to conventional instructional methods (Wouters, van Nimwegen, van Oostendorp, & van Der Spek, 2013). In the serious game, students were made familiar with different means of payment, how to use them and the risks and costs involved.[4] As financial literacy is not part of the education programme in Flanders, the majority of teachers and students had no experience on the content of the course.[5] Students worked together in pairs according to the experimental condition the school was assigned to (see below). We incentivised the students to perform well by providing a small reward for the pair who finished the serious game first (Cerasoli, Nicklin, & Ford, 2014). Teachers were provided a detailed lesson guide. They were given the possibility to use a computerized or paper version of the game [6] and to teach in a block of four classes, two blocks of two classes or four classes separately. All students received paper information booklets while completing the serious game. The booklets were related to topics covered in the serious game and provided further additional information. To ensure the paper and computerized versions to be compatible, the serious game was divided into three modules. Using the paper version, teachers were required to check the solutions of students after each module before handing over the next module. In the computerized version, solutions were checked automatically. If students answered incorrectly, they were requested to revise the material in the booklets once more in order to find the correct solution. As part of the experiment, the role of the teacher was deliberately minimised as teachers were only requested to give a short whole-class introduction on the purpose of the course and instructions on how to solve the serious game. Hence, using the ICT-tool for the majority of classrooms, we reduced the role of the teacher, such that potential additional teacher effects on effectiveness of the financial education programme were minimised.

Group formation. Two forms of grouping are standard in the literature to handle differences in classrooms (Deunk et al., 2015). Students were working in pairs in the financial education programme, either in heterogeneous or homogeneous groups. On the one hand, in a heterogeneous group formation, students' pairs were formed by means of a puzzle game at the start of the course, ensuring a random pair formation and accordingly, student pairs with different abilities. On the other hand, in a homogeneous group formation, we differentiated based on students' cognitive differences related to performance in mathematics. In particular, teachers were asked before the start of the course to form pairs of students according to their performance in mathematics in the previous semester. PISA 2015 results show that

[4] The serious game included exercises on several concepts, i.e. bank accounts, bankcards, bank notes and coins, direct debit and standing orders, skimming, phishing, reliability of information, and the calculation of discounts and costs.

[5] It should be noted that financial literacy compentences will become part of the education programme in Flanders from September 2019 onwards.

[6] Data from the eighth grade shows 10.7 per cent of teachers to implement the paper version in their classroom.

when students do well in mathematics, they also tend to score high in financial literacy, i.e. in Flanders, performance in mathematics and reading is found to explain more than 70% of the variation in financial literacy performance (OECD, 2017). Hence, high ability pairs for which homogeneous group formation was implemented, were expected to do well in the financial education course.

Differentiated instruction. Adapting the didactic material to the needs of students is considered a straightforward approach to differentiate (Kulik, 1992). In the financial education programme, students either received uniform minimal instruction or differentiated instruction. For differentiated instruction, three different versions of the material were designed, all of them resulting in the same set of learning outcomes. A first version consisted of minimal instructions for high ability pairs, a second version for medium ability pairs provided additional instructions in the first module of the course and a third version for low ability pairs additional instructions in all modules of the course. Accordingly, medium and low ability students were given additional hints and cues, such as where to find the answer to a question in the information booklet or how to make a calculation, which enabled to adaptively coach their learning process. Including differentiated instruction may either result in convergence or divergence of students' performance level in a classroom.[7] A priori, we cannot impose that high ability students were sufficiently challenged to find solutions in a self-regulated way. Hence, the dispersion of students' test scores will be examined in the analysis in order to determine the type of differentiation. Further, note that, as all instructions were provided in the course material itself, student pairs were able to work at their own pace.

Treatment conditions. Although the financial literacy programme was very similar to all students, we implemented small variations in the didactic approach, corresponding to the different experimental conditions. In particular, the variation stems from the formation of the students' pairs and the level of instruction given to these students' pairs, as discussed above. In a first treatment condition, schools received the didactic material that involved heterogeneous group formation and no differentiated instruction. In a second experimental condition, the same financial literacy programme was used, but the students were grouped in homogeneous pairs and no differentiated instruction was provided to these pairs. In a third condition, homogeneous group formation was combined with differentiated instruction.

2.2. Survey Data

We assess the impact of the financial education programme on student financial proficiency using test instruments. A computer-aided multiple-choice test was developed by our research team in cooperation with secondary school teachers. Nine questions referred directly to the didactic material on means of payment, while four questions were standard to the financial literacy literature and covered the topics of interest and inflation (Lusardi, Mitchell, & Curto, 2010). The pre-treatment financial literacy test consisted of a short survey on demographics and financial attitudes as well as ten questions on financial

[7] Teachers aiming at convergence wish all students in the classrooms to reach a minimum level of performance, whereas divergence refers to helping all children reach their highest potential (Deunk et al., 2015).

knowledge and three questions on financial behaviour. Except for the rephrasing of some sentences and adjustments of numbers, the post-treatment test questions were designed in a similar way measuring the same concepts. For each test, the duration was equal to 20 minutes on average and was administered in the classroom under supervision of the teacher prior and after the course on means of payment. We derived four outcome measures based on the test instruments, i.e. the course financial proficiency, the financial knowledge, behaviour and attitude. The course financial proficiency combines the financial knowledge and behaviour measures. Students' financial knowledge is measured by means of six questions related to different means of payment, reliability of information, and the calculation of discounts and costs, whereas financial behaviour is assessed by means of three questions related to students' view on being careful with different means of payment and information. The financial attitude is measured using three Likert-scale questions on the importance of financial literacy and saving and whether the student compares prices of shops before making any purchases.

We assess the behaviour of students and teachers by classroom observations. After emails were sent to schools for consent, three researchers examined programme implementation of 17 classes in the first wave and four classes in the second wave using an observation checklist. Accordingly, the findings may provide evidence whether teachers effectively followed the instructions as advised, i.e. whether the effect of the programme itself was not at risk of being mitigated.

2.3. Internal validity

Table II presents the baseline statistics of school and student background and financial characteristics for the final sample of treatment and control schools of the first wave. Due to miscommunication issues, a small fraction of schools received the instructions and course material from another experimental condition, which is one explanation for the difference in the number of observations across conditions. However, it is important to note that this error occurred randomly and that schools were not able to self-select into a specific experimental condition. In addition, in the second wave, we assigned proportionally more schools to the experimental conditions that were underrepresented in the first wave. In line with Bruhn et al. (2016), differences in means are calculated by regressing the characteristic on the treatment variable with standard errors clustered at class level.

Overall, students in treatment and control schools appear to be relatively similar, except for some variables for one of the conditions, i.e. the age, the language spoken at home and the number of holidays per year (which we use as an indicator for student's socioeconomic status). The variable *ability* is measured by the average of student's grades in language (Dutch) and mathematics. Note that this measure only provides a proxy for student's ability as both grades are self-assessed by students on a five-point scale and the grades depend largely on the school and type of education the student is attending. Next, for financial characteristics, baseline differences compared to the control condition can be found as well. Moreover, although the majority of eighth grade students appear to value financial literacy (Likert rating of 4.21 out of 5) and saving (4.17 out of 5) and regularly discuss financial matters

at home (3.13 out of 5), students from treatment and control schools obtain an average grade of 36 per cent. Treatment schools implementing homogeneous group formation with differentiated instruction even perform significantly worse than control schools on the course-related and financial knowledge questions. Further differences related to school variables across treatments can be observed, i.e. for private and public schools, the type of education [8] and class size. The Flemish education system is organised in three educational networks, i.e. GO! education (official education organised by the Flemish community), government-aided public education (run by municipal or provincial authorities) and government-aided private education (organised by a private person or organisation, consisting primarily of catholic schools). Accordingly, our sample mainly consists of government-aided private schools, which is in line with the general representation of the Flemish education system.

Table II: Baseline characteristics Wave I

Variables		**Control**	**Heterogeneous**	*p-value*	**Homogeneous**	*p-value*	**Homogeneous & instruction**	*p-value*
Number of Schools		13	9		9		12	
Number of Students		739	400		349		408	
Type of school (private)		0.68	0.98	0.000	0.92	0.008	0.89	0.017
Class size		17.41 (4.83)	14.52 (4.57)	0.006	15.85 (4.73)	0.126	14.55 (5.12)	0.009
Education type	*General*	653 (88.4%)	360 (90.0%)		306 (87.68%)		261 (63.97%)	
	Technical	21 (2.84%)	32 (6.20%)	0.341	14 (4.01%)	0.668	95 (23.28%)	0.003
	Vocational	65 (8.80%)	8 (1.55%)		29 (8.31%)		52 (12.75%)	
Background characteristics								
Gender (female)		0.54	0.59	0.195	0.45	0.226	0.50	0.503
Age (years)		14.15 (0.44)	14.11 (0.43)	0.251	14.12 (0.36)	0.364	14.26 (0.53)	0.052
Number of holidays per year (4)		2.92 (1.02)	3.12 (1.02)	0.019	2.90 (1.03)	0.797	2.91 (1.02)	0.959
Language (Dutch)		0.84	0.75	0.123	0.91	0.050	0.64	0.003
Ability		3.32 (1.04)	3.19 (1.03)	0.313	3.43 (0.86)	0.361	3.12 (1.09)	0.169
Financial characteristics & scores								
Importance of financial literacy (5)		4.15 (0.75)	4.18 (0.72)	0.688	4.26 (0.65)	0.040	4.05 (0.84)	0.177
Importance of saving (5)		4.23 (0.95)	4.20 (0.91)	0.655	4.34 (0.85)	0.103	4.11 (1.09)	0.120
Comparison shops before purchase (5)		3.60 (1.31)	3.56 (1.35)	0.729	3.77 (1.27)	0.033	3.68 (1.32)	0.326
Discussion of financial matters (5)		3.15 (1.13)	3.05 (1.15)	0.213	3.09 (1.16)	0.575	3.03 (1.18)	0.197
Course financial proficiency (9)		2.91 (1.44)	3.01 (1.56)	0.530	3.20 (1.53)	0.081	2.96 (1.58)	0.791
General financial proficiency (4)		1.74 (0.96)	1.71 (0.98)	0.720	1.63 (0.98)	0.227	1.71 (0.97)	0.651
Financial knowledge (6)		1.78 (1.06)	1.91 (1.16)	0.185	2.03 (1.17)	0.049	1.93 (1.19)	0.209
Financial behaviour (3)		1.13 (0.82)	1.10 (0.76)	0.605	1.17 (0.80)	0.631	1.03 (0.79)	0.211

Baseline statistics of school and student characteristics for the final sample of the second wave are presented in Table III. Again, similar differences between treatment and control schools can be found, i.e. the class size and type of education significantly differ. Across the different conditions, students differ by gender, age, number of holidays per year and how important they perceive financial literacy. Ninth grade students from all treatment and control schools perform poorly on the baseline financial proficiency test as well (with an average grade of 38 per cent) and again, we find schools implementing

[8] In Flanders, eight graders can choose between general and pre-vocational education. Hence, for consistency reasons, we derive their type of education (general/technical/vocational) based on a database of study tracks. Onderwijskiezer, SO: Eerste graad – tweede leerjaar (available at https://www.onderwijskiezer.be/v2/secundair/sec_1graad_2lj.php).

homogeneous group formation with differentiated instruction to perform significantly worse on course-related and financial knowledge questions than control schools.

Table III: Baseline characteristics Wave II

Variables		**Control**	**Heterogeneous**	*p-value*	**Homogeneous**	*p-value*	**Homogeneous & instruction**	*p-value*
Number of Schools		5	7		10		4	
Number of Students		111	161		148		91	
Type of school (private)		0.92	0.95	0.731	0.92	1.00	0.82	0.534
Class size		16.86 (3.63)	10.88 (5.55)	0.007	15.31 (6.28)	0.521	9.35 (3.60)	0.000
Education type	*General*	88 (79.28%)	149 (92.55%)		106 (71.62%)		15 (16.48%)	
	Technical	14 (12.61%)	7 (4.35%)	0.432	23 (15.54%)	0.752	19 (20.88%)	0.046
	Vocational	9 (8.11%)	5 (3.11%)		19 (12.84%)		57 (62.64%)	
Background characteristics								
Gender (female)		0.64	0.60	0.464	0.51	0.038	0.56	0.364
Age (years)		15.14 (0.46)	15.07 (0.43)	0.422	15.26 (0.54)	0.384	15.43 (0.56)	0.026
Number holidays per year (4)		3.03 (1.00)	3.04 (1.03)	0.897	2.97 (0.96)	0.655	2.65 (1.06)	0.037
Language (Dutch)		0.81	0.90	0.261	0.85	0.649	0.69	0.191
Ability		3.09 (0.69)	3.28 (1.00)	0.320	3.18 (0.80)	0.458	3.33 (1.02)	0.247
Financial characteristics & scores								
Importance of financial literacy (5)		4.34 (0.63)	4.30 (0.70)	0.579	4.29 (0.68)	0.506	4.09 (0.80)	0.025
Importance of saving (5)		4.06 (1.05)	4.20 (0.99)	0.410	4.23 (0.95)	0.316	3.98 (1.00)	0.646
Comparison shops before purchase (5)		3.62 (1.33)	3.68 (1.23)	0.612	3.43 (1.29)	0.226	3.41 (1.34)	0.195
Discussion of financial matters (5)		3.23 (1.16)	3.24 (1.12)	0.988	3.22 (1.22)	0.936	3.01 (1.14)	0.150
Course financial proficiency (9)		3.55 (1.71)	3.99 (1.41)	0.157	3.60 (0.95)	0.470	2.56 (1.79)	0.025
General financial proficiency (4)		2.00 (1.00)	1.99 (1.05)	0.972	3.47 (1.47)	0.812	1.65 (1.13)	0.149
Financial knowledge (6)		2.23 (1.25)	2.48 (1.10)	0.270	2.20 (1.15)	0.878	1.51 (1.25)	0.011
Financial behaviour (3)		1.32 (0.82)	1.51 (0.78)	0.129	1.27 (0.84)	0.773	1.05 (0.87)	0.184

As a result of the randomisation procedure, students of treatment and control schools are expected to be equal in expectation, both in observed and unobserved characteristics. However, differences might still occur by chance. Accordingly, the above mentioned school and student characteristics will be accounted for in subsequent analyses.

Next, differences at baseline may also result from the attrition of classes and students in both studies. First, from 334 classes which were signed up by teachers or school principals, 134 (40,12 per cent) classes did not participate in the programme (79 (34,80 per cent) in wave I and 55 (51.10 per cent) in wave II). This relatively high rate of attrition can partially be explained on practical grounds, i.e. as it was possible to implement the programme as a cross-curricular course, some teachers may had to prioritize their own course material over the financial education programme, particularly in the second wave, as it was implemented closer to the end of the academic year.

Second, out of 3,369 students who took the pre-treatment financial proficiency test, 962 (28,55 per cent) students did not complete the post-test (714 (27.36 per cent) in wave I and 248 (32,67 per cent) in wave II). If we restrict to entire classes, i.e. excluding individual students in a class not taking the post-test, 583 (17,30 per cent) students did not comply (407 (15,59 per cent) in wave I and 176 (23,19 per cent) in wave II). See Figure A1 and A2 for a separate presentation of wave I and II. As baseline characteristics were collected for all students, we are able to test for selective attrition. Results show that

classes from schools assigned to the experimental conditions are equally likely to comply as those from control schools, except for the condition of homogeneous group formation with differentiated instruction. Comparing non-complying students across treatment and control schools reveals almost no significant differences. However, when we compare the baseline characteristics of complying and non-complying students, differences can be observed, i.e. students from public schools, those who fail a year in school, with lower socio-economic status and vocational education have a higher probability not completing the post-treatment test (see table AI). Note that this finding may suggest the existence of upper bound effects in subsequent regression analyses.

3. Empirical strategy

3.1. Main analysis

As a result of the randomised nature of the evaluation, a straightforward analysis can be performed. We use the following intention-to-treat OLS regression model:

$$y_{ijs}^{1} = \alpha + \beta_0' treatment_{is} + \beta_{ijs} y_{ijs}^{0} + \sum \beta_1' S_s + \sum \beta_2' X_i + \epsilon_{ij} \quad (1)$$

where y_{ijs}^{1} represents the value of an outcome measure, i.e. the course financial proficiency, financial knowledge, behaviour or attitude for student *i* in class *j* from school *s*, $treatment_{is}$ is a categorical variable indicating to which experimental or control condition student *i* from school *s* is assigned; y_{ijs}^{0} is the pre-treatment value of an outcome for student *i* in class *j* from school *s*; S_s refers to school characteristics for school *s*; X_i indicates individual characteristics (such as the age, gender and SES [9]) of student *i* and is included to control for the baseline imbalances*;* $\epsilon_{i,j}$ is a clustered error term at class level *j* for child *i*. Despite the convention of clustering at the most aggregate level feasible, class-level clustering seems reasonable in our study as the programme implementation and hence, peer-effects occurred within classrooms (Abadie, Athey, Imbens, & Wooldridge, 2017). Moreover, subsequent analysis will show clustering at the level of the school to result in similar estimates. All test measures are standardized and hence, regression coefficients can be interpreted in terms of standard deviations.

3.2. Robustness checks

The coefficients β_0' represent the intention-to-treat effects, i.e. the effects of assigning students to experimental conditions on the outcome measures. Yet, they do not necessarily represent the effects of the financial education programme for those students who actually completed it as instructed. As the data shows a non-compliance rate of 13,5% in the experimental conditions, we account for it using an instrumental variable (IV) approach, estimating the treatment-on-the-treated effects (Athey & Imbens, 2016). Note that the compliance is measured through teacher surveys, administered after the course, in which teachers were asked whether they correctly grouped students and/or gave differentiated instruction as advised. As not all teachers completed the survey, the analysis is limited to a subsample. The dummy variable indicating the treatment assignment serves as an instrument for compliance. The second-stage equation may be presented as follows:

$$y_{ijs}^{1} = \alpha + \beta_0' compliance_{ij} + \beta_{ijs} y_{ijs}^{0} + \sum \beta_1' S_s + \sum \beta_2' X_i + \epsilon_{ij} \quad (2)$$

where $compliance_{ij}$ is determined as a dummy indicating whether the teacher followed the instructions correctly, i.e. grouped students and/or gave differentiated instruction as advised.

[9] Composite indicator of the number of holidays per year and language spoken at home

Next, in order to examine whether the differences across the experimental conditions and control condition found at baseline are important, we perform a Difference-in-Difference (DiD) analysis. The DiD analysis estimates the impact of the financial education programme by netting out the change in each outcome measure between the pre- and post-treatment period among control schools for the change among treatment schools. The regression model is defined as follows:

$$y_{ijst} = \alpha + \beta_0' post_{it} * treatment_{is} + \beta_1' treatment_{is} + \beta_2 post_{it} + \epsilon_{ij} \quad (3)$$

where y_{ijst} represents the pre- and post-treatment values of an outcome measure. The variable $post_{it}$ takes value one for post-treatment observations (zero otherwise). The coefficients of the categorical interaction term $post_{it} * treatment_{is}$ account for the different treatment effects.

4. Results

4.1. Main analysis

Despite the difference in grades, the coefficients for the course financial proficiency, financial knowledge and behaviour of students obtained in the first and second wave are not significantly different from each other, as demonstrated by a Chi-square test in the bottom of Table IV. Positive effects of the financial education programme for all three outcome measures are found in both waves, yet not statistically significant in wave II due to the smaller sample size. Accordingly, these similar results for different age groups may suggest that our findings can be generalized to other settings. Note that, for subsequent analysis, (more precise) pooled estimates will be used.

Table IV: Intention-to-Treat Analysis

Dependent variable		**Course Financial Proficiency**		**Financial Knowledge**		**Financial Behaviour**	
Wave I & II (N = 2,407)	Heterogeneous	0.227***	0.177**	0.254***	0.206***	0.115*	0.0671
		(0.0704)	(0.0702)	(0.0741)	(0.0757)	(0.0627)	(0.0600)
	Homogeneous	0.305***	0.281***	0.284***	0.267***	0.220***	0.189***
		(0.0889)	(0.0768)	(0.0861)	(0.0753)	(0.0784)	(0.0689)
	Homogeneous & instructions	0.126	0.270***	0.107	0.226***	0.0906	0.225***
		(0.0808)	(0.0741)	(0.0815)	(0.0770)	(0.0760)	(0.0683)
Wave I (N = 1,896)	Heterogeneous	0.212***	0.154*	0.222***	0.152*	0.118	0.0896
		(0.0805)	(0.0825)	(0.0837)	(0.0880)	(0.0736)	(0.0726)
	Homogeneous	0.359***	0.303***	0.339***	0.284***	0.252***	0.210***
		(0.107)	(0.0902)	(0.102)	(0.0875)	(0.0904)	(0.0804)
	Homogeneous & instructions	0.152*	0.248***	0.133	0.194**	0.129	0.238***
		(0.0820)	(0.0763)	(0.0813)	(0.0795)	(0.0829)	(0.0775)
Wave II (N = 511)	Heterogeneous	0.256*	0.209	0.346**	0.297*	0.0781	0.0336
		(0.145)	(0.143)	(0.159)	(0.150)	(0.101)	(0.103)
	Homogeneous	0.181	0.210	0.166	0.201	0.109	0.120
		(0.158)	(0.142)	(0.163)	(0.139)	(0.144)	(0.135)
	Homogeneous & instructions	0.0240	0.409*	-0.00963	0.329	-0.101	0.305
		(0.256)	(0.219)	(0.273)	(0.211)	(0.181)	(0.183)
	Controls		X		X		X
Chi-square test (p-value)	Heterogeneous		0.12 (0.731)		0.71 (0.399)		0.20 (0.651)
	Homogeneous		0.32 (0.543)		0.27 (0.606)		0.34 (0.560)
	Homogeneous & instructions		0.50 (0.479)		0.37 (0.541)		0.13 (0.728)

Standard errors clustered at class level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Reference category: control condition; Variables controlled for: gender, SES, type of education, type of school, (grade for full sample); Significant differences across treatment coefficients for Wave I by means of F-test: heterogeneous & homogeneous and instructions for financial behaviour (p = 0.063); Wave II: / ; Wave I & II: heterogeneous & homogeneous (p = 0.086) and heterogeneous & homogeneous and instructions (p = 0.031) for financial behaviour.

When we compare test scores for all three experimental conditions with the control condition, the financial education programme is shown to be effective. In terms of magnitude, after controlling for the variables discussed in section 2, following the financial education programme increases the performance of students in a financial literacy test, ranging from 0.18 to 0.28 standard deviations, on average, compared to receiving no course material. The post-treatment performance of students is primarily

determined by the improvement in financial knowledge (increase of 0.21 to 0.27 standard deviations on average). Financial behaviour of students improves, ranging from 0.07 to 0.23 standard deviations, on average. The results are robust against the inclusion of other control variables (such as financial characteristics) for which imbalance at baseline was found. Further, although we must carefully interpret the results due to the smaller sample size, treatment schools are also found to perform better in the second post-treatment financial literacy test administered as an homework approximately 6 weeks after the course, as represented in table V. [10] Next, we were only able to measure the financial attitude in the post-treatment financial literacy test for ninth grade students in the second wave. Table AII presents the regression analysis with financial attitudes as outcome variables and indicates no improvement of the financial attitude after the programme.

Table V: Second post-treatment test

Dependent variable	**Course Financial Proficiency**		**Financial Knowledge**		**Financial Behaviour**	
Heterogeneous	0.419***	0.450***	0.385***	0.444***	0.253	0.231
	(0.143)	(0.153)	(0.132)	(0.126)	(0.155)	(0.169)
Homogeneous	0.233*	0.206*	0.245**	0.245**	0.166	0.0992
	(0.121)	(0.107)	(0.112)	(0.100)	(0.131)	(0.114)
Homogeneous & instructions	0.270*	0.515***	0.330**	0.590***	0.0565	0.182*
	(0.158)	(0.145)	(0.157)	(0.153)	(0.113)	(0.100)
Controls		X		X		X
Observations	1,208	1,208	1,208	1,208	1,208	1,208

Standard errors clustered at class level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Variables controlled for: gender, SES, type of education, type of school, grade, time between tests; Significant differences across treatment coefficients: heterogeneous & homogeneous (p = 0.086) and homogeneous & homogeneous and instructions (p = 0.07; p = 0.046) for course financial proficiency and financial knowledge, respectively.

When contrasting the three experimental conditions, different group formations and differentiated instruction appear, on average, unimportant for teaching financial education. This finding is in line with previous meta-analyses on group formations (Lou et al., 1996; Slavin, 1990). It holds for the course financial proficiency measure and, in particular, the financial knowledge of students. Per contra, for financial behaviour, significant differences across group formations can be found, i.e. homogeneous group formation significantly increases the financial behaviour measure by 0.13 standard deviations on average in comparison to heterogeneous group formation. Adding differentiated instruction, on the other hand, does not seem to increase the effectiveness of homogeneous group formation. It is worth noting, however, that students in the condition of heterogeneous group formation do not even perform better in the financial behaviour measure compared to students receiving no course material. A more detailed examination of the financial behavior measure reveals that for one of the three questions used in the measure, students in both the control condition and the condition of heterogeneous group formation seem to perform worse in the post-treatment test. One potential explanation for this matter is the rephrasing of the particular multiple-choice question in the post-treatment test for which the solution

[10] Although only 23 per cent of students completed the second post-treatment test due to the homework format, the comparison of complying and non-complying students reveals no significant differences pointing at potential upper or lower bound effects.

could only be found if students had carefully read the information booklet. Accordingly, students in heterogenous groups may have overlooked it, whereas in the other conditions they did not.

Although our analysis indicates no general effects of the different group formations and differentiated instruction on the course financial proficiency measure, it is important to examine whether this holds for *all* students. As shown in Figure AIII, the performance of students in our study improved across the distribution of post-treatment course financial proficiency scores, which is represented by a rightward shift in the scores of treatment schools compared to control schools. However, despite that the differences in distributions are statistically significant, as measured by the Kolmogorov-Smirnov test, we can easily observe the gains in students' performance to be less pronounced at the bottom and top of the distribution of post-treatment course financial proficiency scores, in particular for the condition of homogeneous group formation with differentiated instruction.[11] Moreover, despite the obtained estimates being rather imprecise, quantile regressions for each decile of the course financial proficiency measure provide some insights in the heterogeneity of treatment effects. As significant differences throughout the deciles can be found, group formations and differentiated instruction seem to have differential impacts on students with different abilities (see Table AIII and Figure AIV).

Table VI: Heterogeneous effects by means of subgroup analysis

Dependent variable: Course financial proficiency	**Ability**			**Language at home**		**Class size**	
	Low	***Medium***	***High***	***Dutch***	***Other***	***Small average***	***Large***
Heterogeneous	-0.00128	0.190*	0.259**	0.178**	0.220*	0.287***	-0.0691
	(0.110)	(0.101)	(0.103)	(0.0758)	(0.117)	(0.0877)	(0.112)
Homogeneous	0.0859	0.273***	0.347***	0.259***	0.368**	0.273***	0.297**
	(0.137)	(0.103)	(0.0932)	(0.0768)	(0.176)	(0.102)	(0.130)
Homogeneous & instructions	0.212*	0.247**	0.312***	0.178**	0.478***	0.346***	0.374***
	(0.119)	(0.114)	(0.108)	(0.0808)	(0.122)	(0.0910)	(0.0965)
Controls	X	X	X	X	X	X	X
Observations (N)	400	1,208	799	1,922	485	1,336	801

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Reference category: control condition; Variables controlled for: gender, SES, type of education, type of school, grade; Significant differences across coefficients by F test for low ability: heterogeneous & homogeneous grouping with differentiated instructions (p = 0.055); for other language: heterogeneous & homogeneous grouping with differentiated instructions (p = 0.014); for large class: heterogeneous & homogeneous grouping with differentiated instructions (p = 0.000) and heterogeneous & homogeneous grouping (p = 0.008)

The prevailing heterogeneity is examined more in detail with the following subgroup analyses. First, we compare students actually eligible or not for additional instructions. As low ability students received additional instructions in all modules, we attempt to identify these students by means of their overall grade in language and mathematics. That is, whenever this grade is 60 per cent or below, the student will have received additional instructions. A medium ability student is suggested to have an overall grade above 60 percent, but less than 70 per cent. We identify high ability students if they have an

[11] The Kolmogorov-Smirnov test examines the equality of distributions. In our sample, the distribution of post-treatment course financial proficiency scores of each experimental condition are compared with the control condition. The tests rejects equality of distribution for all three experimental conditions (p = 0.000, p = 0.000 and p = 0.076).

overall grade of 70 per cent or more. Hence, these students received minimal instructions. Table VI shows that the performance of low ability students significantly improves by 0.21 standard deviations compared to receiving no course material whenever they are grouped homogeneously and given additional instructions, in contract to hetero- or homogeneous group formation without additional instructions.[12] Medium and high ability students, on the other hand, do not seem to perform significantly different in any of the experimental conditions.

Next, the result also holds when comparing students who speak Dutch (the language in which the financial education programme was taught) or another language at home. Again, we find that students speaking another language at home significantly benefit from differentiated instruction, while others do equally well in hetero- or homogeneous groups with or without differentiated instruction.

Finally, the class size can also be considered as a mediating factor in the effectiveness of group formations and differentiated instruction. Accordingly, the subgroup analysis for large classes illustrates homogeneous group formation to be more efficient than heterogeneous group formation, yet, adding differentiated instructions does not increase the effectiveness. For students in small classes, different group formations and differentiated instruction seem unimportant within the context of financial education.

4.2. Robustness analysis

We can examine whether our results are robust against the following approaches. First, it is worth noting that the estimates of table IV do not indicate whether the different teaching delivery methods of the financial education programme were effective for those students whom teacher implemented the programme as instructed – treatment-on-the-treated effects. By means of an IV regression analysis, we estimate the treatment-on-the-treated effects on a subset of the sample, as presented in table AIV in appendix. Despite, as expected, the coefficient sizes going up (in particular for the condition of homogeneous group formation), the interpretation of our main results remains valid, i.e. no differences can be found between hetero- and homogeneous group formations. Differentiating instruction does not seem to improve the student performance for the latter group formation.

A second robustness test implies clustering standard errors at the school level. Although clustering at class level, the level where peer effects occurred, seemed more reasonable in our analysis, table AIV indicates the results of clustering standard errors at the more conservative school level to be consistent with those obtained when clustering at class level.

Third, the Difference-in-Differences analysis indicates whether differences in characteristics found at baseline are important. Except for the financial knowledge measure, the obtained results in table AIV appear relatively similar to our main intention-to-treat results which suggests our sample to be balanced well.

[12] The result is driven by the performance on the financial behaviour questions.

5. Discussion

The impact of financial education programmes on the financial proficiency of secondary school students has been reported before. Consistent with our findings, Bruhn et al. (2016) found the overall performance of students to increase by a quarter of a standard deviation, while other experimental studies indicate more modest improvements, around 0.15 standard deviations (Frisancho, 2018; Villanueva et al., 2018). Note, however, as section 2.3 on internal validity indicated the presence of potential upper bound effects, our estimates may be biased upwards to some extent, which is a matter that can be examined further by means of a bounds analysis for selective attrition.

We find that the improvement of financial literacy is driven by an increase in financial knowledge. The financial behaviour, in particular for the treatment of heterogeneous group formation, and the financial attitudes are more moderately or unaffected by the financial education programme. One potential explanation for the lack of effects on financial attitudes is that the target population in our study was too young and hence, had limited exposure to money. Nonetheless, our data reveals the majority of students to save and compare prices before purchasing at baseline (see table I and II). Alternatively, the data we have to measure financial attitude might be too limited, including only 504 observations for ninth grade students. Consequently, an evaluation using more observations may result in larger significant effects, such as in Bruhn et al. (2016).

On the whole class level, taking together low, medium and high ability students, no general effects of the different group formations and differentiated instruction on the course financial proficiency and financial knowledge measures are observed. For financial behaviour, on the other hand, we find homogeneous group formation to be more effective than heterogeneous group formation, yet, differentiated instruction does not improve performance. Our analysis does show distinct heterogeneity of treatment effects, which is in line with the education literature on differentiation (e.g. Deunk et al., 2015; Lou et al., 1996; Slavin, 1990). Different patterns for students with different abilities and backgrounds are found. Students with low ability or speaking another language at home significantly benefit from a homogeneous group formation within the classroom, given the material includes additional instructions. The performance of students with medium or high performance or students speaking Dutch at home, on the other hand, remains similar in all experimental conditions. This finding has an important implication, i.e. it indicates that providing differentiated instruction for struggling students does not come at the expense of other students when groups of students are formed homogeneously within the classroom. As the performance of all student with different ability levels improves, we may conclude that high ability students are sufficiently challenged as well by the minimal instructions in order to find solutions in a self-regulated way. Accordingly, homogeneous group formation combined with differentiated instruction enables to help *all* students reach their highest potential. This result primarily holds for the financial behaviour of students.

Students in large classrooms are expected to significantly benefit from differentiated instruction compared to students in small classrooms. Apart from the larger likelihood of heterogeneity in student performances within the classroom, teachers in large classrooms need to divide their individual instruction time among a larger number of students as well. Although we observe homogeneous group formation to be more efficient than heterogeneous group formation, adding differentiated instruction does not increase the effectiveness, which is in contrast to previous research (Barrow, Markman, & Rouse, 2009). Per contra, Kulik & Kulik (1982) demonstrated that group formations generally have no impact on student achievement, yet, a positive influence on students' attitude during class. The classroom observations reveal this finding to hold in our study as well, i.e. the observation checklists indicate that in classrooms where teachers implemented homogeneous group formation, students were more likely to collaborate with their peer, were more actively involved and less distracted.

A final point to consider is why the effects of differentiation instruction are not as distinct as expected. One potential explanation is that, although the programme was designed in such way that the role of teachers was minimized, teachers still got involved with their students. When teachers do not follow instructions as advised, e.g. help their students with the content of the course, the effect of differentiated instruction implemented in the course material itself may have been mitigated. Examining this matter qualitatively, classroom observations indeed indicate these events to have occurred in a subgroup of classrooms. Not only does this potentially explain the lack of effects for differentiated instruction in general and for students in larger classrooms, it may also indicate that the effects for low ability students are likely to be underestimated. A limitation of our study, however, is that we cannot analyse this matter empirically.[13]

[13] Including class or teacher fixed effects in the regression analyses omits all treatment coefficients. The classroom observation data remains too limited to perform empirical analyses.

6. Conclusion

Significant heterogeneity prevails in today's classrooms, and even more pronounced in financial literacy education. Up to now, randomised evaluations in the financial literacy field have provided uniform financial education programmes without accounting for the student heterogeneity in the proficiency level or ability. This study conducted two waves of randomised experiments measuring the impact of different teaching delivery methods of a financial education programme in eighth and ninth grade classrooms of Flemish secondary schools. As new insights with respect to the importance of the learning context and financial literacy delivery methods were provided, our findings contribute to the literature on economics of education in general, and financial literacy education in particular.

In order to solve for the prevailing heterogeneity of students' performance in financial literacy, differentiation practices were integrated in a financial education programme with the aim of helping all students reach their highest potential within classrooms. Accordingly, the effectiveness of heterogeneous and homogeneous group formations were examined along with the added-value of differentiating instruction to students' ability, as measured by their performance in mathematics, in a financial education programme. The analysis combined the two experimental waves and examined the financial proficiency of students on several dimensions, i.e. the financial knowledge, behaviour and attitude. Our analysis indicated the financial education programme to be effective on two dimensions. Overall, financial knowledge and behaviour of students significantly improved, yet, financial attitudes remained unaffected. Further, our results showed significant heterogeneity in treatment effects when comparing the different teaching delivery methods. Homogeneous group formation combined with differentiated instruction seemed to be an important differentiation tool. Accordingly, students with lower ability significantly improved their performance in a financial literacy test, whereas the performance of medium and high ability students in the classroom was not deteriorated by it.

As teachers are unable to meet the diverse needs of students when teaching *to the middle*, our study provides causal evidence that diverse educational policies on differentiation can effectively be implemented in contemporary classrooms.

Appendix

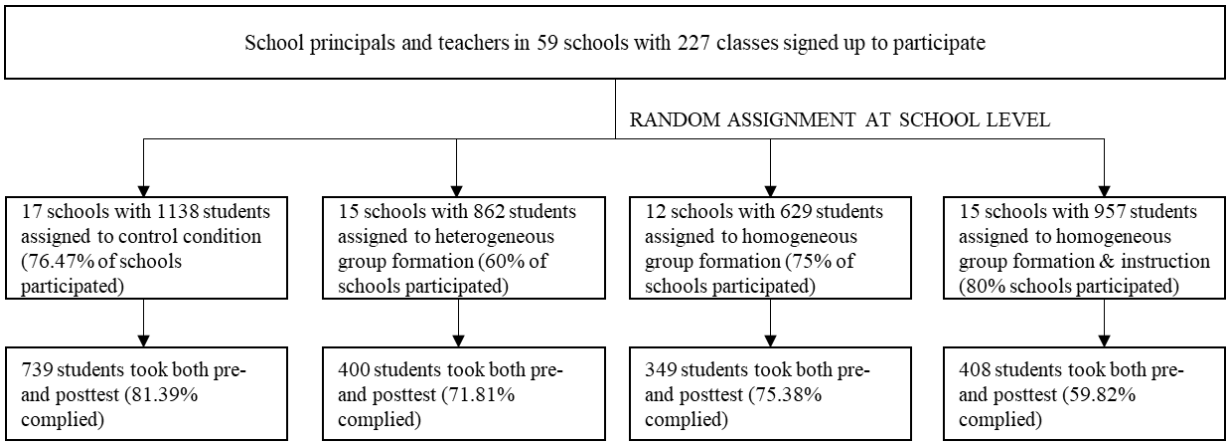

Figure AI: Selection of participants per treatment for Wave I

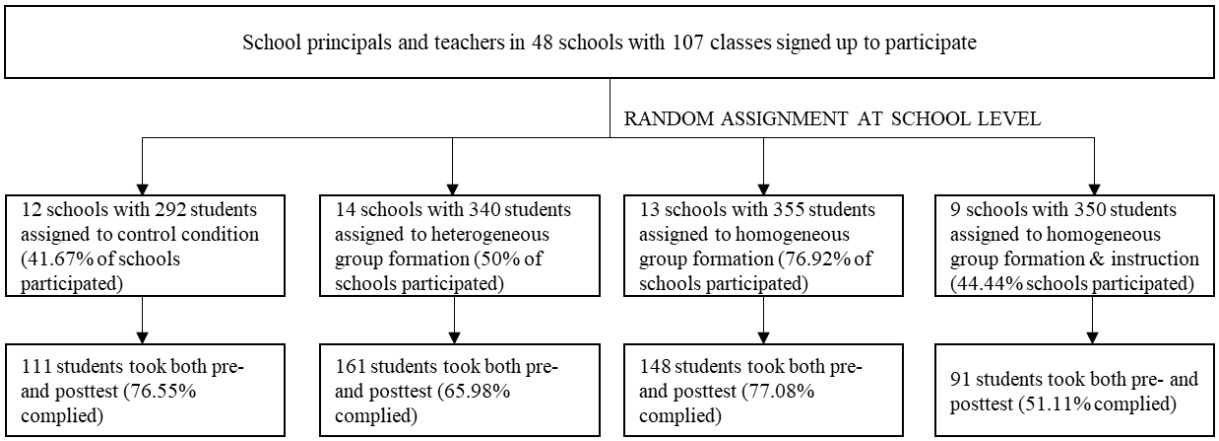

Figure AII: Selection of participants per treatment for Wave II

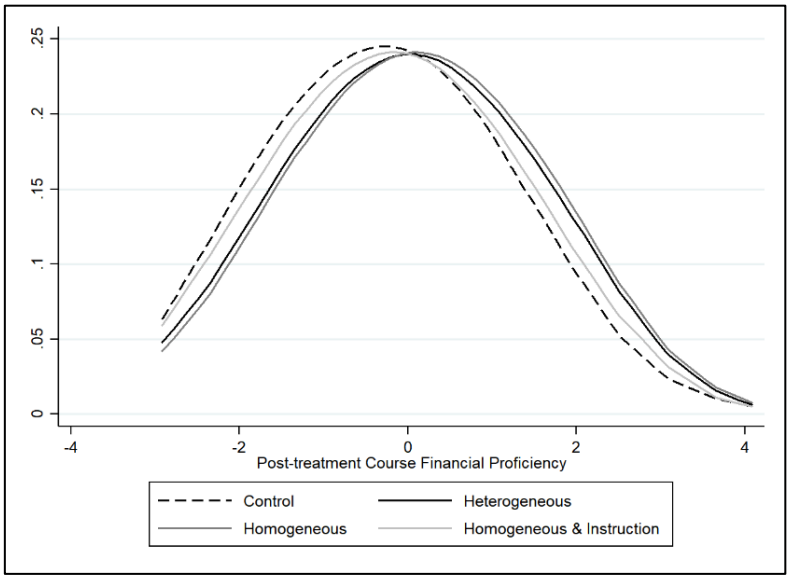

Figure AIII: Distribution Shift of Course Financial Proficiency

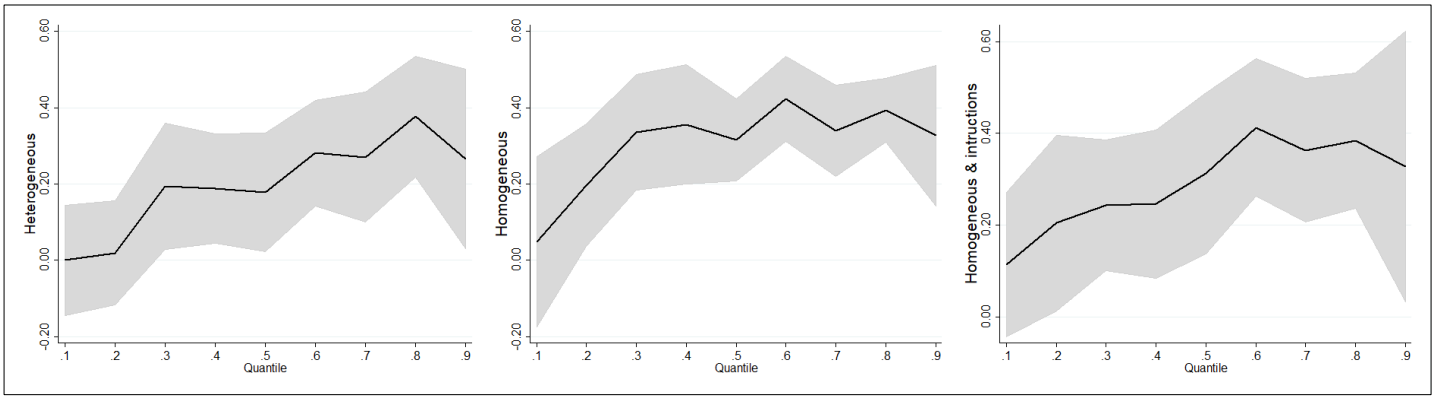

Figure AIV: Quantile regressions for Course Financial Proficiency

Table AI: Selective attrition

Characteristic	Type of school (Private)	Type of education	Gender (female)	Age	Number holidays per year	Language (Dutch)	Ability
Attrition	-0.239**	0.337*	-0.0736	0.186***	-0.159*	-0.0188	0.0451
	(0.121)	(0.177)	(0.0725)	(0.0640)	(0.0881)	(0.0375)	(0.108)
Heterogeneous*Attrition	0.115	0.139	-0.0488	0.0145	-0.139	-0.0440	-0.127
	(0.134)	(0.228)	(0.0872)	(0.0953)	(0.137)	(0.0631)	(0.148)
Homogeneous*Attrition	-0.0530	-0.248	0.0659	-0.0424	0.210	-0.0217	0.00557
	(0.165)	(0.227)	(0.0969)	(0.112)	(0.133)	(0.0583)	(0.185)
Homogeneous & instructions*Attrition	0.0694	-0.200	0.0643	-0.0540	0.0670	-0.0352	-0.335**
	(0.148)	(0.231)	(0.0856)	(0.0993)	(0.142)	(0.0724)	(0.162)

Characteristic	Import. financial literacy	Import. saving	Comparison shops	Discussion financial matters	Course financial prof.	General financial prof.
Attrition	-0.0784	-0.0399	0.135	0.0353	-0.139	-0.0931
	(0.0602)	(0.103)	(0.0909)	(0.0797)	(0.149)	(0.0881)
Heterogeneous*Attrition	-0.0779	-0.119	-0.245*	0.0215	-0.436*	0.0504
	(0.0901)	(0.135)	(0.143)	(0.112)	(0.241)	(0.122)
Homogeneous*Attrition	-0.0205	0.0834	0.0115	-0.345**	0.0973	0.104
	(0.0867)	(0.130)	(0.157)	(0.136)	(0.249)	(0.121)
Homogeneous & instructions*Attrition	0.0920	-0.101	-0.189	0.137	0.0195	0.00801
	(0.0940)	(0.135)	(0.137)	(0.117)	(0.223)	(0.111)
Observations	3,369	3,369	3,369	3,369	3,369	3,369

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table AII: Financial attitude of ninth grade students

Dependent variable	Importance financial literacy		Importance saving		Comparison shops before purchase	
Heterogeneous	-0.309	-0.355	-0.00672	0.0197	-0.143	-0.134
	(0.320)	(0.288)	(0.207)	(0.204)	(0.186)	(0.195)
Homogeneous	0.366	0.369	0.229	0.244	-0.323	-0.318
	(0.297)	(0.276)	(0.181)	(0.181)	(0.244)	(0.246)
Homogeneous & instructions	-0.741***	-0.579*	-0.404	-0.416	-0.566***	-0.365
	(0.288)	(0.312)	(0.317)	(0.349)	(0.193)	(0.288)
Controls		X		X		X
Observations	504	504	504	504	504	504

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; estimates derived from ordered logistic regression; note 4 missing values for financial attitude outcome measures

Table AIII: Quantile regressions for Course Financial Proficiency

Dependent variable:	**Course Financial Proficiency**								
Course financial proficiency	**10th**	**20th**	**30th**	**40th**	**50th**	**60th**	**70th**	**80th**	**90th**
Heterogenous	-0	0.0191	0.194**	0.188***	0.178**	0.281***	0.271***	0.376***	0.266**
	(0.101)	(0.0943)	(0.0810)	(0.0657)	(0.0764)	(0.0678)	(0.0821)	(0.0846)	(0.113)
Homogenous	0.0487	0.197**	0.335***	0.356***	0.315***	0.424***	0.340***	0.394***	0.327***
	(0.106)	(0.0956)	(0.0944)	(0.0949)	(0.0574)	(0.0772)	(0.0812)	(0.107)	(0.119)
Homogenous & intructions	0.114	0.205**	0.243**	0.246***	0.313***	0.412***	0.363***	0.384***	0.327**
	(0.0963)	(0.0904)	(0.0974)	(0.0876)	(0.0931)	(0.0741)	(0.0777)	(0.0969)	(0.128)
Controls	X	X	X	X	X	X	X	X	X
Observations (2,407)									

Bootstrap standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; Variables controlled for: gender, SES, type of education, type of school, grade

Table AIV: Robustness tests

Dependent variable		**Course Financial Proficiency**		**Financial Knowledge**		**Financial Behaviour**	
Test I – *IV analysis*	Heterogeneous	0.290**	0.279**	0.301*	0.293	0.142	0.128
		(0.127)	(0.127)	(0.175)	(0.182)	(0.105)	(0.0902)
	Homogeneous	0.525***	0.445***	0.511***	0.432***	0.325*	0.260*
		(0.158)	(0.128)	(0.141)	(0.116)	(0.169)	(0.146)
	Homogeneous & instructions	0.178**	0.326***	0.178*	0.292***	0.0950	0.243***
		(0.0904)	(0.0804)	(0.0910)	(0.0810)	(0.0860)	(0.0809)
	Controls		X		X		X
	Observations	1,516	1,516	1,516	1,516	1,516	1,516
Test II – *Clustered s.e. at school level*	Heterogeneous	0.227**	0.177*	0.254**	0.206**	0.115**	0.0671
		(0.0923)	(0.0913)	(0.101)	(0.100)	(0.0550)	(0.0551)
	Homogeneous	0.305***	0.281***	0.284**	0.267***	0.220***	0.189***
		(0.107)	(0.0888)	(0.115)	(0.0972)	(0.0736)	(0.0667)
	Homogeneous & instructions	0.126	0.270***	0.107	0.226**	0.0906	0.225***
		(0.106)	(0.0891)	(0.107)	(0.0877)	(0.0853)	(0.0813)
	Controls		X		X		X
	Observations	2,407	2,407	2,407	2,407	2,407	2,407
Test III – *DiD approach*	Heterogeneous	0.136*		0.122		0.0957	
		(0.0745)		(0.0797)		(0.0647)	
	Homogeneous	0.217**		0.146		0.214**	
		(0.104)		(0.111)		(0.0859)	
	Homogeneous & instructions	0.181**		0.102		0.203**	
		(0.0874)		(0.0868)		(0.0905)	
	Controls						
	Observations	4,814		4,814		4,814	

Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Variables controlled for: gender, SES, type of education, type of school, grade; Significant differences across treatment coefficients for test I by means of F test: /; Significant differences across treatment coefficients observed for test II by means of F-test: heterogeneous & homogeneous grouping (p = 0.037) and heterogeneous & homogeneous with differentiated instructions (p = 0.051) for financial behaviour.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). When should you adjust standard errors for clustering? National Bureau of Economic Research. *NBER Working Paper No. 24003*.

Alavania, P., & Farhady, S. (2012). Using Differentiated Instruction to Teach Vocabulary in Mixed Ability Classes with a Focus on Multiple Intelligences and Learning Styles. *International Journal of Applied Science and Technology*, *2*(4), 72–82.

Athey, S., & Imbens, G. (2017). The Econometrics of Randomised Experiments. Handbook of Economic Field Experiments, 1:73-140.

Barrow, L., Markman, L., & Rouse, C. E. (2009). Technology's Edge: The Educational Benefits of Computer-Aided Instruction. *American Economic Journal: Economic Policy*, *1*(1), 52–74.

Baumgartner, T., Lipowski, M. B., & Rush, C. (2003). *Increasing Reading Achievement of Primary and Middle School Students through Differentiated Instruction.* (ERIC No. ED479203).

Becchetti, L., Caiazza, S., & Coviello, D. (2013). Financial education and investment attitudes in high schools: Evidence from a randomised experiment. *Applied Financial Economics*, *23*(10), 817–836.

Becchetti, L., & Pisani, F. (2012). Financial education on secondary school students: The randomised experiment revisited. AICCON Working Papers 98.

Berry, J., Karlan, D., & Pradhan, M. (2018). The Impact of Financial Education for Youth in Ghana. *World Development*, *102*, 71–89.

Betts, J. R., & Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review*, *19*, 1–15.

Bruhn, M., de Souza Leão, L., Legovini, A., Marchetti, R., & Zia, B. (2016). The Impact of High School Financial Education: Evidence from a Large-Scale Evaluation in Brazil. *American Economic Journal: Applied Economics*, *8*(4), 256–295.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, *140*(4), 980–1008.

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

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, *101*(August), 1739–1774.

Figlio, D. N., & Page, M. E. (2002). School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality ? *Journal of Urban Economics*, *51*, 497–514.

Frisancho, V. (2018). The Impact of School-Based Financial Education on High School Students and Their Teachers: Experimental Evidence from Peru. *IDB Working Paper Series No IBD-WP-871*.

Gamoran, A., & Weinstein, M. (1998). Differentiation and Opportunity in Restructured Schools. *American Journal of Education*, *106*(3), 385.

Giorgi, G. De, Pellizzari, M., & Woolston, W. G. (2012). Class size and class heterogeneity. *Journal of the European Economic Association*, *10*(4), 795–830.

Hinojosa, T., Miller, S., Swanlund, A., Hallberg, K., Brown, M., & O 'brien, B. (2010). *The Impact of The Stock Market Game on Financial Literacy and Mathematics Achievement: Results from a National Randomised Controlled Trial*.

Kulik, J. A., & Kulik, C.-L. C. (1991). *Research on ability grouping: Historical and contemporary perspectives*. Storrs: University of Connecticut, National Research Center on the Gifted and Talented. (ERIC Document Reproduction Service No. ED 350 777).

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & D'Apollonia, S. (1996). Within-Class Grouping : A Meta-Analysis. *Review Of Educational Research*, *66*(4), 423–458.

Lührmann, M., Serra-Garcia, M., & Winter, J. (2015). Teaching teenagers in finance: Does it work? *Journal of Banking and Finance*, *54*, 160–164.

Lusardi, A., & Mitchell, O. S. (2014). The Economic importance of Financial Literacy: Theory and Evidence. *Journal Of Economic Literature*, *52*(1), 5–44.

Lusardi, A., Mitchell, O. S., & Curto, V. (2010). Financial Literacy among the Young. *The Journal of Consumer Affairs*, *44*(2), 358–380.

Mandell, L. (2008). *The Financial Literacy of Young American Adults: Results of the 2008 National Jump$trat Coalition Survey of High School Seniors and College Students*. Washington DC.

OECD. (2017). *PISA 2015 Results (Volume IV): Students' Financial Literacy*. Paris.

Simpkins, P. M. C., Mastropieri, M. A., & Scruggs, T. E. (2009). Differentiated Curriculum Enhancements in Inclusive Fifth-grade Science Classes. *Remedial and Special Education*, *30*(5), 300–308.

Slavin, R. E. (1990). Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis. *Review Of Educational Research*, *60*(3), 471–499.

Tieso, C. (2005). The Effects of Grouping Practices and Curricular Adjustments on Achievement. *Journal for the Education of the Gifted*, *29*(1), 60–89.

Tomlinson, C. A. (1999). *The differentiated Classroom Responding to the Needs of All Learners* (first, Vol. 85). Alexandria, Virginia: ASCD.

Tomlinson, C. A. (2001). *How to Differentiate Instruction in Mixed-Ability Classrooms*. Association for Supervised and Curriculum Development.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, *27*(2–3), 119–145.

Valiandes, S. (2015). Evaluating the impact of differentiated instruction on literacy and reading in mixed ability classrooms: Quality and equity dimensions of education effectiveness. *Studies in Educational Evaluation*, *45*, 17–26.

van Rooij, M., Lusardi, A., & Alessie, R. J. M. (2011). *Financial literacy, retirement planning, and household wealth*. CFS Working Paper , No. 2011/2.

Villanueva, E., Bover, O., & Hospido, L. (2018). The Impact of High School Financial Education on Financial Knowledge and Choices : Evidence from a Randomized Trial in Spain. Discussion Paper Series IZA DP No 11625.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265.