The Effect of Group Formation and Differentiated Instruction in Financial Literacy Education. Evidence from Two Randomised Control Trials. [†]

Kaat Iterbeke, Kristof De Witte, Koen Declercq and Wouter Schelfhout[*]

Abstract. The present paper examines the impact of group formation and differentiated instruction on the learning outcomes of eighth and ninth grade students in a financial education programme. In particular, the effect of within-class heterogeneous and homogeneous group formations is investigated by forming pairs of students either randomly or based on their abilities. In addition, the paper studies whether differentiated instruction, with additional instructions for lower ability students, enhances the impact of the financial education programme. The paper provides evidence on effects of both the practices using two randomised control trials involving 69 schools and 2,407 students. Overall, the results suggest that the programme induces higher financial proficiency in the students by almost a quarter of standard deviation. Although the learning outcomes of the average student are not found to be affected by group formation and differentiated instruction, non-native students are seen to significantly benefit from homogeneous group formation and differentiated instruction.

Keywords: Financial Literacy; Group Formation; Differentiated Instruction; Randomised Control Trial.

JEL-classification: C93; I21;O16.

[†] The financial support from the Flemish Science Organisation is acknowledged through the grant S000617N. This study is registered in the AEA RCT Registry and the unique identifying number is AEARCTR-0003478. We are grateful to Els Lagrou and Danièle Van der Espt for research assistance. We thank conference and seminar participants in Brussels, Leuven and Maastricht for valuable suggestions on an earlier version of this paper.

[*] Iterbeke: Leuven Economics of Education Research, KU Leuven, Leuven, Belgium (e-mail: kaat.iterbeke@kuleuven.be); De Witte: Leuven Economics of Education Research, KU Leuven, Leuven, Belgium; UNU-MERIT, Maastricht, the Netherlands (e-mail: kristof.dewitte@kuleuven.be, k.dewitte@maastrichtuniversity.nl); Declercq: IRES, UCLouvain and FNRS, Louvain-la-Neuve, Belgium (e-mail: koen.declercq@uclouvain.be); Schelfhout: Antwerp School of Education, University of Antwerp, Antwerpen, Belgium (e-mail: wouter.schelfhout@uantwerpen.be).

1. Introduction

Nowadays, the classrooms are becoming increasingly diverse. In order to address heterogeneity within classrooms, many countries have implemented tracking in secondary education which involves segregating the students into different educational tracks based on their abilities. School tracking may have negative consequences on learning outcomes in the event of misallocation of students or low mobility between different educational tracks. Hanushek and Woessman (2006) found that early tracking in schools is likely to increase educational inequality. Moreover, even in the countries with tracking in secondary education, differences in learning profiles, cultural and economic backgrounds within classrooms persist such that the teachers are required to adapt to their way of teaching. However, some argue that similar 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; Schiltz et al., 2019).

Using a randomised evaluation, the present paper examines the consequences of educational policies aimed towards dealing with classroom heterogeneity by exploring the effects of two differentiation practices, i.e. within-class group formations and differentiated instruction on student's learning outcomes. From a policy perspective, examining these effects is important as manipulating the class composition and adopting differentiated instruction in order to cope with heterogeneity may increase the effectiveness of education at quite lower costs as compared to other policies such as the ones aimed at class size reductions (Giorgi, Pellizzari, & Woolston, 2012). Further, as tracking is particularly harmful for students from a disadvantaged background (Hanushek & Woessman, 2006), differentiation might avoid the negative impact on these subgroups. The results show that group formation and differentiated instruction do not significantly affect learning outcomes of the average student, but that they are particularly beneficial for non-native students.

Research suggests classroom diversity to be pronounced in financial literacy education (Lusardi & Mitchell, 2014). This could be attributed to the prevailing multiplicity in the opportunities available to improve financial literacy. Such opportunities are majorly linked to gender, socioeconomic status and immigrant background (OECD, 2017). Financial literacy is considered to be an essential skill for strengthening one's presence in today's society due to the increasingly complex and challenging financial landscape. On the one hand, financial literacy implies the knowledge of financial concepts and risks, and on the other hand, it indicates the presence of skills and motivation to apply such knowledge (OECD, 2017). Among other factors, high levels of financial literacy are positively correlated with wealth accumulation, retirement planning and debt management (van Rooij, Lusardi, & Alessie, 2011). However, despite its importance, the existing literature established pervasive low levels of financial literacy around the world (Lusardi & Mitchell, 2014). Accordingly, the governments, non-profit organisations, and financial institutions have been promoting the development of financial education

programmes. Although the former remedial approaches targeted merely the adults, the focus has now been shifted to school-based interventions promoting financial literacy. Apart from financial illiteracy being more pronounced among youth, school-based programmes are attractive for several reasons (Lusardi, Mitchell, & Curto, 2010; Mandell, 2008). Such programmes not only ensure financial guidance at the right time, i.e. when the students are still developing habits and have not made major financial decisions as adults, but also simultaneously offer the possibility to minimise selective participation thereby generating spillover effects as the students are likely to act as the agents of change in their family's financial decisions (Frisancho, 2018; Walstad et al., 2017).

The widespread financial illiteracy calls for causal evidence on the effectiveness of financial education initiatives (OECD, 2017). Accordingly, there is a growing body of literature discussing the impact of financial literacy education among youth with some recent experimental studies in favour of financial education programmes establishing that they increased financial knowledge thereby also affecting attitudes and behaviour (e.g., 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, 2018; Villanueva, Bover, & Hospido, 2018).

To the best of our knowledge, the present research is novel in its approach as the previous evaluations on school-based financial education programmes did not consider heterogeneity of students' performances in financial literacy at the planning phase of the programmes, i.e. it was considered that all students were being taught the same programme in a uniform way. Given the increasing heterogeneity in classrooms, a differentiated approach, aimed at addressing the particular needs of *all* students, may be more effective. A set of hallmarks for effective differentiation are proposed in the educational science literature, among which, the use of small flexible groups in the classroom is one such approach (Tomlinson et al., 2003). It is believed that the small-group settings are found to give teachers the flexibility to address student variance more appropriately than they are able to while instructing the whole class. However, meta-analyses that compared homogeneous and heterogeneous group formations, by holding methods of instructions as constant, revealed little or no significant differences (Lou et al., 1996; Slavin, 1990).[1] Yet, different performance patterns for students with different abilities within classrooms have been observed with negative effects for low ability students in homogeneous groups, whereas positive effects for medium ability students and no difference in either group formation for high ability students (Deunk, Doolaard, Smale-Jacobse, & Bosker, 2015). Hence, a second approach which is found to be effective is the adaptation of instruction and material to suit the needs of all groups (Lou et al., 1996). The materials should explicit instructional needs of the students, such as giving additional support to low ability students and more challenging assignments to high achievers (Slavin, 1990). From

[1] 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.

the few studies examining the effect of computer-assisted instruction on direct student outcomes and employing a credible identification strategy, the literature showcases mixed evidence (Banerjee, Duflo, & Linden, 2007; Barrow, Markman, & Rouse, 2009; Rouse & Krueger, 2004). As the effectiveness differed by the country and study domain, i.e. more support was found in mathematics instruction rather than reading, it is unclear how these results are to be generalized to other domains and settings.

Our paper contributes to the literature by evaluating the causal impact of differentiated instruction, and homogeneous and heterogeneous group formations within a class on learning outcomes. In particular, the paper contributes to the experimental literature on financial literacy by evaluating whether these differentiation practices enhance the effectiveness of a school-based financial education programme. New insights concerning the importance of the learning context and financial literacy delivery methods are provided.

Two large-scale randomised control trials were set up in the eighth and ninth grade of secondary school in the northern region of Belgium, Flanders to establish the results of the paper. Examining the different age groups not only enables to observe age-specific heterogeneity, it can also be used to assess the validity of results, i.e. if the results appear to be similar in both studies, the likelihood of generalization to other contexts increases. The financial education programme was offered as a serious game (game-based learning) for which students were made to work in pairs. By means of a random assignment of schools to specific conditions, the paper distinguishes the impact of different group formations and differentiated instruction from the financial education programme in general without differentiation measures. Schools were either assigned to a control condition, receiving no financial literacy education, or one out of three experimental conditions. In the first experimental condition, all students were paired randomly and received the same material. To analyse the impact of a homogeneous group formation, the students were paired according to their abilities in the second experimental condition. Finally, to investigate the effectiveness of differentiated instruction, the paper considered the same homogeneous group formation as in the previous experimental condition, but additional instructions were provided to lower ability students in the third experimental condition.

The findings of our paper suggests that the financial education programme is effective, i.e. the programme is seen to increase the financial proficiency of students by almost a quarter of a standard deviation. Although no major effects of group formations and differentiated instruction are found, strong heterogeneous effects are observed as non-native students benefit from homogeneous group formation and differentiated instruction. Moreover, suggestive evidence is provided that the performance of low ability students improves when a homogeneous group formation in the classroom is implemented and additional instructions are provided, while scores of high ability students, who do receive minimal instructions, still increases relatively more in the particular setting.

The remaining paper unfolds as follows. In the next section, the secondary education system and financial literacy in Flanders will be discussed. Section 3 will describe the design and the data.

Section 4 will present the empirical strategy and corresponding results. In section 5, some potential mechanisms will be discussed along with the conclusion.

2. Secondary education and financial literacy in Flanders

This section reviews the relevant characteristics of secondary education and the average level of financial literacy in Flanders. The Flemish education system is organised into three educational networks, i.e. 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. It should be noted that the majority of Flemish schools, and accordingly our sample, falls into the latter category. Despite all networks receive similar government funding, private schools attract, on average, students with a higher socioeconomic status. Flemish secondary education is organized in a tracking system. In particular, the eighth graders have the liberty of choosing from elective courses within an academic or pre-vocational education track. Students in ninth grade are tracked in an academic, technical, arts or vocational education track.

Although students in Flanders perform above average in the PISA financial literacy assessment, distinct heterogeneity in students' performances is observed within the region (OECD, 2017). Moreover, financial literacy was not part of the education curriculum in Flanders at the time of intervention. As a consequence, teachers and students had no experience on the content of the material taken for this paper.[2]

3. Experimental design

Two randomised control trials involving secondary school students in eighth and ninth grade in the region of Flanders were conducted. The final sample includes 2,407 students in 200 classes in 69 schools. The first experimental study took place between January, 2018 and March, 2018. The sample for this first study consists of 1,896 students in 148 eighth grade classes in 43 schools. The second experimental study was implemented in ninth grade and covers data on 511 students in 52 classes in 26 schools. The second study started in February, 2018 and ended in June, 2018.

In what follows, an overview of the study timeline, and the content and implementation of the financial education programme are given. Next, the differentiation practices and differences between experimental conditions are elaborated. Finally, details on the test instruments and outcome measures are provided.

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

3.1. Procedure

Secondary schools were invited in an open call to use novel didactic material on financial literacy. The schools that signed up were randomly assigned to a control condition and multiple experimental conditions. Randomising at school level was done to ensure minimization of contamination and spillover effects for schools with multiple classes. Moreover, to guarantee a uniform implementation of the material across all conditions, the teachers were requested to follow a strict protocol. Participation in the intervention was conditioned on the acceptance of this protocol. First, to avoid interference of parents, it was decided that the material had to be delivered during regular class hours. Second, to measure the value added by the intervention, all students were made to undergo 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. The first post-treatment test had to be administered immediately after the completion of the material, whereas the second post-treatment test was administered as a homework approximately six weeks after the lectures. Finally, to avoid timing effects, schools were informed to deliver the material in pre-specified periods. Table I presents the study timeline for treatment and control schools of study I and II.

Table I: Study timeline for study I and II

STUDY I	**Jan'18**	**Mid-End Jan'18**	**End Jan'18**		**End Mar'18**	**May'18**
Treatment schools	Randomization to an experimental condition	Student pre-test (*2 weeks to complete)*	Material sent to schools	*8 weeks to teach 4 lectures*	1st post-test & teacher survey	2nd post-test
Control schools	Randomization to control condition	Student pre-test (*2 weeks to complete)*	/	/	1st post-test & teacher survey	/
STUDY II	**End Feb'18**	**End Feb-Mar'18**	**Mid Apr'18**		**Jun'18**	**End Jun'18**
Treatment schools	Randomization to an experimental condition	Student pre-test (*2 weeks to complete)*	Material sent to schools	*6 weeks to teach 4 lectures*	1st post-test & teacher survey	2nd post-test
Control schools	Randomization to control condition	Student pre-test (*2 weeks to complete)*	/	/	1st post-test & teacher survey	/

Note: Teachers from treatment schools of study II were given 6 weeks instead of 8 to plan the lectures due to time restrictions and exam periods.

The Financial Education Programme. To examine the impact of within-class group formations and differentiated instruction in financial literacy education, four conditions were designed; one control condition in which schools did not receive the didactic material, and three experimental conditions. All experimental conditions made use of very similar material, which were designed by senior teachers in the research team in the form of four lectures of 50 minutes each on the topic 'means of payment' in the format of a computerised serious game. This format was used in order to create an interactive environment which directed to a clear goal and provided immediate feedback to students. Apart from the practical implementations, serious games are found to be more effective in terms of learning and

retention as compared to the conventional instructional methods (Wouters, van Nimwegen, van Oostendorp, & van Der Spek, 2013). Also Hinojosa et al. (2010) used a game format in their financial education programme and found significant positive treatment effects.

During the lectures, students were made familiar with different means of payment, how to use them and the risks and costs involved.[3] This topic is part of one of the four content areas of the PISA financial literacy assessment, 'money and transactions' (OECD, 2016). Students worked together in pairs and had to solve several questions on the topic. To solve these questions, all students were given paper information booklets while completing the material. The booklets were related to topics covered in the material and provided certain additional information. The material was divided into three modules and students were able to start the next module only when they correctly answered all questions of the previous module. If students answered incorrectly, they were requested to revise information in the booklets once more in order to find the correct solution. The students were motivated to perform well by extending a small reward for the pair who successfully finished the material first. To guarantee a uniform implementation in all participating schools, the teachers received clear instructions which explained the material as a deliberately autonomous task for students in which the teacher should intervene as little as possible. In addition, the material led students independently through the different steps, making a teacher intervention uncalled for. The teachers were only requested to give a short introduction to the entire class explaining the purpose of the course and instructions on how to navigate through the material. Hence, by using the ICT-tool for the majority of classrooms, the role of the teacher was reduced, such that potential additional teacher effects on effectiveness of the financial education programme were minimised by design.[4]

Group formation. The literature suggests two forms of group formations to handle differences in classrooms (Deunk et al., 2015). On the one hand, in a heterogeneous group formation, pairs of students were formed through a puzzle game at the start of the course to ensure that such pairs were formed randomly. On the other hand, in a homogeneous group formation, the students were differentiated based on their ability. In particular, before the start of the course, teachers were requested to form pairs of students according to their performance in mathematics in the previous semester. PISA 2015 results show that the students performing well in mathematics also tend to score high in financial literacy. For example, in Flanders, performance in mathematics and reading is found to explain more than 70 percent of the variation in financial literacy performance (OECD, 2017). Hence, it was expected that the high ability pairs for which homogeneous group formation was implemented would perform better during the lectures.

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

[4] If schools lack proper ICT infrastructure, they could use a paper-based version of the material. Data from the eighth grade shows 10.7 percent of teachers to implement the paper version in their classroom. Results from a regression analysis indicate no difference in the effectiveness of the computerized or paper-based version of the didactic material.

Differentiated instruction. In the financial education programme, student pairs worked at their own pace and were either given uniform minimal instruction or differentiated instruction according to their ability level. For differentiated instruction, three different versions of the material were designed and all of them resulted in the same set of learning outcomes. The first version consisted of minimal instructions for high ability pairs. The second version, for medium ability pairs, provided additional instructions in the first module of the material, while the third version, for low ability pairs, included further instructions in all the modules of the material. 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 them to coach their learning process adaptively (see figure AI). Including differentiated instruction may either result in convergence or divergence of performances in a classroom.[5] It cannot be asserted a priori that high ability students were challenged sufficiently to find solutions in a self-regulated way and that they benefited from the differentiated instruction. Hence, the heterogeneity of students' test scores will be examined in the subsequent analysis in order to determine whether the differentiated instruction resulted in convergence or divergence of the test scores.

Experimental conditions. Although the financial literacy programme was similar for all the students, variations in the didactic approach corresponding to the different experimental conditions were implemented. In particular, the variation stems from the division of students into pairs and the level of instruction given to these pairs of students, as discussed above. In the first experimental condition, a random selection of schools were given the material that involved heterogeneous group formation and no differentiated instruction. In the second experimental condition, the same financial literacy programme was used, but the students were grouped into homogeneous pairs and no differentiated instruction was provided to these pairs. Comparing students in these two experimental conditions allows us to identify the impact of group formation on learning outcomes. In the third condition, homogeneous group formation was used along with differentiated instruction to evaluate whether differentiated instruction enhances the impact of the programme in a homogeneous group formation. Table II presents an overview of the different experimental conditions.

Table II: Control and experimental conditions

	Control condition	Experimental condition I **Heterogeneous**	Experimental condition II **Homogeneous**	Experimental condition III **Homogeneous & instructions**
Group formation	No	Random	Based on ability	Based on ability
Instruction	No	Uniform	Uniform	Differentiated

Note: In what follows, the paper denotes the different experimental conditions as indicated in bold in the table.

[5] 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).

3.2. Test instruments

The impact of the financial education programme on student financial proficiency was assessed using test instruments. A computer-aided multiple-choice test consisted of nine questions which led directly to the material on means of payment. The pre-treatment financial literacy test consisted of a short survey on demographics and statements related to financial attitudes, six questions on financial knowledge and three questions on financial behaviour. The post-treatment test used the similar questions measuring the same concepts with minor rephrasing and number adjustments (see Appendix AIII for detailed description). For each test, the duration was eleven minutes on an average, and it was conducted in the classroom under the supervision of the teacher before and after the completion of the material.

Three outcome measures are derived based on these test instruments, i.e. the 'course financial proficiency', the financial knowledge and behaviour. The 'course financial proficiency' measure combines the financial knowledge and behaviour measures. Students' financial knowledge is measured through six questions related to different means of payment, reliability of information, and the calculation of discounts and costs (questions 1, 2, 3, 4, 5, and 7 in Appendix AIII). Financial behaviour is assessed through three questions related to students' view on being careful with different means of payment and information (questions 6, 8, and 9 in Appendix AIII).

The behaviour of students and teachers was assessed by classroom observations. After emails were sent to schools for consent, three experienced PhD researchers were made to examine programme implementation of 17 classes in the first study and four classes in the second study using a standardised observation checklist. When allowed by school regulations, the classes were filmed. The classroom observations provide evidence on the exact implementation in the schools.

4. Data

This section describes the collected data which will be used in the subsequent regression analyses, and the size and importance of attrition observed in both studies.

4.1. Descriptive statistics

Table III presents the descriptive statistics for the composite sample (the descriptive statistics for the two experimental studies separately are presented in Table AI). In total, 2,407 pupils in 69 schools were randomly allocated to one of the four conditions. 72 percent of students in the control condition attend private schools, whereas the corresponding shares are higher in all experimental conditions. Next, most students are enrolled in the academic track, yet this share is lowest in the experimental condition of homogeneous group formation combined with differentiated instruction. To test for differences in school and pupil characteristics between the control and experimental conditions, a regression of the characteristic on the treatment indicator is estimated with clustered standard errors at class level. Overall, the students in treatment and control schools appear to be relatively similar, however, some significant differences can be observed.

The next panel of Table III shows the background characteristics of pupils. The sample contains slightly more girls than boys and the students are on average 14 years old. The proxy for the socioeconomic status is a composite indicator consisting of the language spoken at home and the number of times a student has travelled abroad with a maximum value of four referring to students who travelled more than two times abroad during the last year. We use this proxy as students easily recall how often they have travelled abroad and this measure is expected to be correlated with household income. We find most students to speak Dutch at home (the official language of the region of Flanders), (the official language of the region of Flanders), however, in the third experimental condition this share is rather low. Finally, the variable *ability* is measured by the grade point average (GPA) in language (Dutch) and mathematics. This measure merely provides a proxy for the ability of students as the GPA is self-assessed on a five-point scale and the grades largely depend on the school and track the student is acquiring (i.e., there are no standardised tests in the Flemish education system).

Table III: Descriptive Statistics Study I & II

Variables	**Control**	**Heterogeneous**	*p*-value	**Homogeneous**	*p*-value	**Homogeneous & instruction**	*p*-value
Number of Schools	17	16		17		15	
Number of Students	850	561		497		499	
Type of school (private)	0.72	0.97	0.000	0.92	0.007	0.88	0.036
Class size	17.34 (4.69)	13.47 (5.14)	0.000	15.85 (4.73)	0.095	13.60 (5.27)	0.000
Track *Academic*	741 (87.2%)	509 (90.73%)		412 (82.90%)		276 (55.31%)	
Technical	35 (4.12%)	39 (6.95%)	0.214	37 (7.44%)	0.646	117 (22.85%)	0.002
Vocational	74 (8.71%)	13 (2.32%)		48 (9.66%)		109 (21.84%)	
Background characteristics							
Gender (female)	0.55	0.59	0.257	0.47	0.139	0.51	0.415
Age (years)	14.28 (0.56)	14.38 (0.61)	0.239	14.46 (0.67)	0.104	14.47 (0.70)	0.045
Number of holidays per year (4)	2.93 (1.02)	3.09 (1.02)	0.018	2.92 (1.01)	0.827	2.87 (1.03)	0.401
Language (Dutch)	0.84	0.79	0.310	0.89	0.095	0.65	0.001
Ability	3.29 (1.01)	3.22 (1.02)	0.508	3.35 (0.85)	0.528	3.16 (1.08)	0.294
Financial characteristics							
Importance of financial literacy (5)	4.18 (0.74)	4.21 (0.71)	0.515	4.27 (0.66)	0.051	4.06 (0.83)	0.067
Importance of saving (5)	4.21 (0.97)	4.20 (0.93)	0.884	4.31 (0.88)	0.103	4.09 (1.07)	0.085
Comparison shops before purchase (5)	3.60 (1.32)	3.56 (1.32)	0.983	3.66 (1.28)	0.407	3.63 (1.33)	0.700
Pre-treatment financial scores							
Course financial proficiency (9)	3.00 (1.49)	3.29 (1.58)	0.048	3.28 (1.51)	0.054	2.89 (1.63)	0.488
Financial knowledge (6)	1.84 (1.10)	2.07 (1.17)	0.021	2.08 (1.16)	0.031	1.85 (1.21)	0.915
Financial behaviour (3)	1.16 (0.82)	1.21 (0.79)	0.375	1.20 (0.81)	0.530	1.03 (0.81)	0.107
1st post-treatment financial scores							
Course financial proficiency (9)	2.97 (1.65)	3.52 (1.78)	0.001	3.65 (1.72)	0.000	3.16 (1.72)	0.247
Financial knowledge (6)	1.94 (1.21)	2.34 (1.25)	0.000	2.38 (1.26)	0.000	2.09 (1.24)	0.189
Financial behaviour (3)	1.02 (0.89	1.18 (0.94)	0.039	1.27 (0.94)	0.003	1.07 (0.93)	0.546

Note: Mean values and standard deviations in parentheses. For the track, absolute and relative frequencies are given. The p-values are derived from difference in means test between students in treatment and control schools (by means of a regression of the characteristic on the treatment indicator with clustered standard errors at class level). Financial scores are reported before standardization.

The next two panels of Table III report the financial literacy of students as measured at baseline. Although the majority of students appear to value financial literacy (mean score of 4.18 out of 5) and saving (4.20 out of 5), and compare prices of shops before purchasing (3.61 out of 5) students in treatment and control schools obtain an average score of 3.12 out of 9. At baseline, the schools implementing heterogeneous or homogeneous group formation perform better on the course financial proficiency and financial knowledge measure than the control schools.

The final panel of Table III reports the results on the post-treatment test and provides a first assessment of the effectiveness of the financial literacy programme. After the financial education programme, students in treatment schools, in contrast to control schools, improve and obtain higher scores than observed at baseline. Hence, the significant differences compared to control schools give a preliminary indication of positive treatment effects. Note that these first results should be interpreted with caution given the differences in student characteristics and initial financial literacy differences between the several experimental conditions. Similar differences are found for both experimental studies separately in Table AI in appendix.

4.2. Attrition

There could be various reasons for the attrition in the present experiments. First, from 334 classes which were registered by teachers after the open call, 134 (40.12 percent) classes did not participate in the experiment (79 (34.80 percent) in study I and 55 (51.40 percent) in study II). The relatively high rate of attrition observed here can partially be explained on practical grounds, i.e. as there had been a possibility of implementation of the programme as a cross-curricular course, some teachers might have had to prioritize their own course material over the financial education programme, particularly in the second study, as it was implemented closer to the end of the school year. Second, from the 3,369 students who took the pre-treatment financial proficiency test, 962 (28.55 percent) students did not complete the post-test (714 (27.36 percent) in study I and 248 (32.67 percent) in study II). Students might not have completed the post-test because of absenteeism or because their teacher failed to adhere to the protocol. See Figure AII and AIII for a separate presentation of study I and II.

Baseline characteristics of all students who completed the pre-test (both compliers and non-compliers) are presented in Table AII for both studies. Although randomisation typically guarantees observed and unobserved characteristics to be equal at this stage, similar differences as in Tables III are observed. This finding can be explained by the high rate of attrition of schools prior to the completion of pre-test. Therefore, in order to assess the external validity of our results and to compare non-complying and complying schools, characteristics of three different sets of schools are compared by matching school-level data on socioeconomic indicators (the percentage children with a mother having at most a secondary education degree, the percentage children not speaking the official language (Dutch) at home, and the percentage children receiving an allowance) to unique school identifiers participating

in the study.[6] First, Table AIII shows the sample of 107 schools that were willing to participate in the programme to be representative of the average Flemish school. Second, the comparison of complying schools (final sample of schools) and the total population of Flemish schools indicates no significant differences in terms of the socioeconomic composition either. When the complying and the non-complying schools are compared, significant differences for two socioeconomic indicators (although only at ten per cent level) are observed. Students in non-complying schools are more likely to receive an allowance and have a mother with at most a secondary education degree.

Next, as the baseline characteristics were collected for all students, it is possible to test for selective attrition between the pre- and post-test. This is done by regressing an indicator for having missing data for the outcome measures on treatment indicators and controls. Table AIV shows that students from schools assigned to the conditions of heterogeneous group formation and homogeneous group formation combined with differentiated instruction were more likely to drop out in the first study. The finding remains significant when combining both studies. As teachers and students had to follow very similar protocols in all experimental conditions, and as we do not observe this in the second study, it is believed that this pattern of attrition was not caused by design. The comparison of non-complying (only completing the pre-test) and complying students (completing the post-test as well) within each condition indicates significant differences for both the studies, as represented in Table AV. In particular, for the control condition and condition of heterogeneous group formation, non-complying students are found to be more likely, among others, to follow technical or vocational tracks, to repeat a grade, to have lower socioeconomic status, and moreover, to perform worse in the pre-treatment financial literacy test. Accordingly, this analysis of attrition demonstrates the importance of including control variables and assessing the robustness of our regression results (e.g., by matching techniques and bounding the estimates).

5. Results

In the following subsections, the empirical strategies and corresponding regression analyses are discussed. Further, through an Instrumental Variable analysis, Lee bounds, and Coarsened Exact Matching, the robustness of the main results of this paper with respect to compliance issues, attrition and covariate imbalances is assessed.

5.1. Intent-To-Treat analysis

As a result of the randomised nature of the evaluation, a straightforward analysis can be performed. The following intent-to-treat (ITT) OLS regression model is used:

[6] AGODI, Cijfermateriaal - Leerlingenkenmerken (2017-2018), available at http://www.agodi.be/cijfermateriaal-leerlingen kenmerken.

$$y_{ijs}^{1} = \alpha + \beta_0' treatment_{is} + \beta_1 y_{ijs}^{0} + \sum \beta_2' S_s + \sum \beta_3' X_i + \epsilon_{ij} \quad (1)$$

where y_{ijs}^{1} represents the value of an outcome measure, i.e. the course financial proficiency, financial knowledge or behaviour for student *i* in class *j* in school *s*; $treatment_{is}$ represents a categorical variable indicating to which experimental or control condition student *i* in school *s* is assigned; y_{ijs}^{0} is the pre-treatment value of an outcome for student *i* in class *j* in school *s*; S_s refers to school characteristics (such as the type of school) for school *s*; X_i indicates individual characteristics (such as the gender, SES, and track) of student *i* and is included to control for the baseline imbalances*;* ϵ_{ij} is a clustered error term at class level *j* for student *i*. Clustering was done at class level as the programme implementation and, hence, potential peer effects occurred within classrooms (Abadie, Athey, Imbens, & Wooldridge, 2017). Moreover, a robustness analysis with standard errors clustered at school level suggests similar significance levels (see table AVIII). All test measures are standardized and hence, regression coefficients can be interpreted in terms of standard deviations.

Table IV: Intent-to-Treat Analysis

Dependent variable		**Course Financial Proficiency**		**Financial Knowledge**		**Financial Behaviour**	
Study 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)
	Controls	No	Yes	No	Yes	No	Yes
Study 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)
	Controls	No	Yes	No	Yes	No	Yes
Study 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	No	Yes	No	Yes	No	Yes
Chi-square test (*p*-value)	Heterogeneous		0.12		0.71		0.20
			(0.731)		(0.399)		(0.651)
	Homogeneous		0.32		0.27		0.34
			(0.543)		(0.606)		(0.560)
	Homogeneous & instructions		0.50		0.37		0.13
			(0.479)		(0.541)		(0.728)

Note: Standard errors clustered at class level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Pre-treatment value of outcome is controlled for in all specifications. Controls: gender, SES, track, type of school, (grade when Study I & II combined); Significant differences across treatment coefficients calculated by means of F-tests - Study I: Heterogeneous and Homogeneous & instructions (p = 0.063) for financial behaviour measure - Study II: / - Study I & II: Heterogeneous and Homogeneous (p = 0.086) and Heterogeneous and Homogeneous & instructions (p = 0.031) for financial behaviour measure. The Chi-square test statistic is calculated to examine whether the treatment coefficients significantly differ between Study I and Study II.

The three first panels of Table IV show the intent-to-treat estimates for the composite sample and each experimental study separately. For each outcome measure, the regression in the first column includes the treatment indicator and the pre-treatment value of the outcome, while the regression in the second column additionally controls for the variables discussed above. Despite the age differences, the coefficients for the course financial proficiency, financial knowledge and behaviour of students obtained in the first (panel 2 in Table IV) and second study (panel 3 in Table IV) are not seen to be significantly different from each other, as demonstrated by a Chi-square test in panel 4 of Table IV. Positive effects of the financial education programme for all three outcome measures are found in both studies, yet not statistically significant in study II due to the smaller sample size. Accordingly, the similar results for different age groups in the two experimental studies may suggest that our findings can be generalized to other settings. Given the similarities and to converge the discussion, in subsequent analysis, the (more precise) pooled estimates of the composite sample are discussed.

Comparing the outcomes of the experimental conditions to those of the control condition, the results suggest that the financial education programme is effective. In terms of magnitude, controlling for observed heterogeneity, following the financial education programme increases the performance of students in a financial literacy test by almost a quarter of a standard deviation, on average. Intuitively, assuming a normal distribution, a quarter of a standard deviation is the difference between a student at the 40th percentile and the 50th percentile. The estimated coefficients range from 0.18 to 0.28 standard deviation. The primary observation concerning the post-treatment performance of students is the improvement in financial knowledge (increase of 0.21 to 0.27 standard deviation, on average). Financial behaviour of students improves from 0.07 to 0.23 standard deviation, on average.[7] All 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 which implemented a homogeneous group formation and differentiated instruction are also found to perform better in the second post-treatment financial literacy test, as represented in table AVII.[8]

Contrasting the three experimental conditions, different group formations and differentiated instruction appear, on average, unimportant for financial literacy education. This finding is in line with previous meta-analyses on group formations (Lou et al., 1996; Slavin, 1990). It holds true for the course financial proficiency measure and, in particular, the financial knowledge of students. Per contra, for financial behaviour, significant differences across group formations are found, i.e. a homogeneous group

[7] The financial attitude of ninth grade students is measured using three Likert-scale statements based on the importance of financial literacy and saving, and the likelihood of the student comparing prices of shops before making any purchases. As the post-treatment measurement of financial attitudes was only done in the second study, results are reported in appendix. Table AVI suggests no improvement of students' financial attitude. The age of the target population, which is suggested to have limited exposure to money, and the limited sample size may explain the absence of significant impacts.

[8] Although only 23 percent of students in treatment schools completed the second post-treatment test due to the homework format, the comparison of complying and non-complying students reveals no significant differences. Students from control schools did not complete the second post-treatment test. Accordingly, their first post-treatment test scores are used as a comparison, considering no learning effects of the test itself.

formation significantly increases the financial behaviour measure by 0.13 standard deviations on average in comparison to heterogeneous group formation. Adding differentiated instruction does not seem to increase the effectiveness of homogeneous group formation. It is worth noting, however, that students in the heterogeneous condition do not perform better on the financial behaviour measure when compared to the students who received no financial literacy education.

5.2. Heterogeneous effects

Despite the absence of effects of group formations and differentiated instruction on average test scores, it is important to examine whether this holds for *all* students. As shown in Figure AIV, the performance of students improved across the distribution of post-treatment course financial proficiency scores, which is represented by a rightward shift in the scores of students in the experimental conditions compared to students in the control condition. However, as measured by the Kolmogorov-Smirnov test, differences in distributions are statistically significant, yet, it is observed that the gains in students' performance are less pronounced at the bottom and top of the distribution of post-treatment course financial proficiency scores. In particular, this holds true for the condition of homogeneous group formation with differentiated instruction.[9] Moreover, quantile regressions for each decile of the course financial proficiency measure provide more detailed insights in the heterogeneity of treatment effects. Not only is the programme seen as effective for all deciles, differences can also be found throughout the deciles. Group formations and differentiated instruction seem to have differential impacts on students with different proficiency levels as, for example, students in the lower deciles seem to benefit from differentiated instruction, whereas students in the highest decile improve similarly in either experimental condition (see Figure AV).

The prevailing heterogeneity is examined comprehensively using the following subgroup analyses. First, the students who were (in)eligible for additional instructions are compared. Hence, low, medium, and high ability students are identified by means of the self-assessed grade in mathematics. That is, a third of the students within each classroom with the lowest grade in mathematics is assumed to have low ability. High ability students are identified as a third of the classroom population with the highest grade in mathematics. Medium ability students are identified as the remaining students left in the classroom. Table V shows the effect size for low ability students, who were eligible for additional instruction in all modules of the material, to be the greatest (0.24 standard deviation) whenever they are grouped homogeneously and given additional instructions. Although medium and high ability students seem to perform better in the condition of homogeneous group formation without differentiated instruction, scores of high ability students, who did not receive any additional instructions, still improve more (0.31

[9] 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 reject equality of distribution for all three experimental conditions ($p = 0.000$, $p = 0.000$ and $p = 0.076$).

standard deviation) than those of low ability students when differentiated instruction is provided in the classroom.

Next, the students who speak Dutch (the native language) are compared to the ones speaking another language at home. Non-native students are found to 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 while assessing the effectiveness of group formations and differentiated instruction. Apart from the greater likelihood of heterogeneity in student performances within larger classrooms, teachers in large classrooms are also required to divide their individual instruction time among more students. Accordingly, the subgroup analysis for large classes (defined as a class with 19 students or more) presents homogeneous group formation to be more efficient than heterogeneous group formation, yet, it is seen that addition of differentiated instructions does not increase the effectiveness. For students in smaller classes, different group formations and differentiated instruction seem trivial within the context of financial literacy education.

Table V: Heterogeneous effects by means of subgroup analysis

Dependent variable: **Course Financial Proficiency**	**Math ability**			**Language at home**		**Class size**	
	Low	***Medium***	***High***	***Dutch***	***Other***	***Small average***	***Large***
Heterogeneous	0.172	0.132	0.211**	0.178**	0.220*	0.279***	-0.0691
	(0.109)	(0.0979)	(0.105)	(0.0758)	(0.117)	(0.0771)	(0.112)
Homogeneous	0.217**	0.277**	0.325***	0.259***	0.368**	0.285***	0.297**
	(0.109)	(0.110)	(0.0945)	(0.0768)	(0.176)	(0.0890)	(0.130)
Homogeneous & instructions	0.243**	0.220**	0.309***	0.178**	0.478***	0.325***	0.374***
	(0.0994)	(0.110)	(0.109)	(0.0808)	(0.122)	(0.0803)	(0.0965)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	739	799	869	1,922	485	1,606	801

Note: Standard errors clustered at class level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Controls: pre-treatment value of outcome, gender, SES, track, type of school, grade; Significant differences across treatment coefficients calculated by means of F-tests - other language: Heterogeneous and Homogeneous & instructions (p = 0.014) - large class: Heterogeneous and Homogeneous & instructions (0.000) and Heterogeneous and Homogeneous (p = 0.008).

5.3. Robustness analysis

5.3.1. Non-compliance

The coefficients in the previous section represent the effects of the intent-to-treat analysis, 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 percent in the experimental conditions, non-compliance is controlled for using the assigned treatment allocations as instruments for actual implementation of group formation and differentiated instruction such that the local average treatment effects are estimated (Athey & Imbens, 2017). The programme compliance is measured through teacher

surveys, administered after the course, in which teachers were asked whether they correctly grouped the students (either heterogeneous or homogeneous) and gave differentiated instruction depending on the experimental condition the school was assigned to. Since not all the teachers completed the survey, the analysis is limited to a subsample. Because of the experimental set-up, the instruments meet the inclusion and exclusion restrictions. The first and second-stage equations may be presented as follows:

$$1^{st} \text{ stage: } compliance_{ij} = \mu + \gamma_0' treatment_{is} + \gamma_{ijs} y_{ijs}^0 + \sum \gamma_1' S_s + \sum \gamma_2' X_i + \delta_{ij} \quad (2)$$

$$2^{nd} \text{ stage: } y_{ijs}^1 = \alpha + \beta_0' \widehat{compliance}_{ij} + \beta_{ijs} y_{ijs}^0 + \sum \beta_1' S_s + \sum \beta_2' X_i + \epsilon_{ij} \quad (3)$$

where $compliance_{ij}$ denotes a dummy indicating whether the teacher of student *i* in class *j* followed the instructions correctly, i.e. grouped students and/or gave differentiated instruction as advised.

Table AIX shows that, despite the increase in the size of the coefficients (in particular for the condition of homogeneous group formation), the interpretation of the main results remains valid, i.e. no differences are found between the heterogeneous and homogeneous group formations. Moreover, differentiating instructions does not seem to improve the student performance.[10]

5.3.2. Sample attrition and Lee bounds

If students selectively exit the sample, the reported treatment effects might be biased. Therefore, to examine the extent to which the differential attrition could be present in the results of this paper, estimates are bound by using the bounding approach proposed by Lee (2009). Two identifying assumptions are required for Lee bounds, i.e. a random assignment to experimental conditions and a monotonicity condition such that the treatment assignment affects attrition in one direction only. Both assumptions, a priori, hold true in both studies. The upper and lower bounds are generated by trimming the sample of the conditions that are least affected by selection to equalize response rates across conditions. Main intent-to-treat models are then re-estimated on the resulting sample (see Figure AVI for a detailed description of the bounds computation). In order to examine at which particular stage the estimates appear robust, we identify a set of bounds for different portions (from 10 to 100 percent) of the actual trimming fractions, corresponding to less versus more restrictive bounds. Figure AVI, AVII, and AVII in appendix present the Lee bounds for all outcome measures for the composite sample, study I, and study II, respectively.

The figures present that the main intent-to-treat estimates are well within the bounds. In figure AVI, we observe that at a trimming portion of 30 percent of what should be trimmed to equalize response rates across conditions, the confidence intervals of the estimates and the upper and lower bounds for the heterogeneous and homogeneous group formation overlap and hence, these intent-to-treat estimates appear robust at this stage. For the estimate of homogeneous group formation combined with differentiated instruction, this occurs at a trimming portion of 50 percent.

[10] The same interpretation remains valid when excluding the non-complying classes from the intent-to-treat regression analysis.

5.3.3. Matching

An alternative robustness test accounting for confounding influence of covariates is the Coarsened Exact Matching (CEM) approach proposed by Iacus, King, and Porro (2008) and King, Blackwell, Iacus, and Porro (2009). CEM enables generation of covariate balance *ex ante* and the comparison of learning outcomes of similar students in the control and experimental condition. In particular, each covariate is temporarily coarsened into meaningful strata such that an 'exact matching' algorithm is applied to the coarsened data in order to identify matches between the control and experimental condition. After unmatched observations are deleted, the original uncoarsened values of covariates are retained and main intent-to-treat models can be re-estimated on the resulting sample. As CEM is designed for binary treatment variables, the model specification is altered such that all experimental conditions are separately compared to the control condition. Accordingly, main (unbalanced) intent-to-treat models are re-estimated on these samples as well. Examining both regression models in Table AX, the results of the main intent-to-treat analysis prove to be robust after matching, suggesting no confounding influence of covariates.

6. Discussion and conclusion

Significant heterogeneity prevails in today's classrooms which is even more pronounced in financial literacy education. In the previous literature, randomised evaluations have been performed on uniform financial education programmes without accounting for heterogeneity in the proficiency or ability levels of students. This paper conducted two randomised control trials measuring the impact of different teaching delivery methods on the effectiveness of a financial education programme in eighth and ninth grade classrooms in Flemish secondary schools. The paper contributes to the literature on the economics of education in general, and financial literacy education in particular, by providing new insights on the importance of group formation and differentiated instruction.

The results show that financial proficiency improves by a quarter of a standard deviation after the financial education programme. Consistent with our findings, Bruhn et al. (2016) found that the overall performance of students increased by similarly, while other experimental studies indicate more modest improvements, around 0.15 standard deviations (Frisancho, 2018; Villanueva et al., 2018). Further, it was found that the improvement of financial literacy was mainly driven by an increase in financial knowledge. The financial behaviour, especially for the treatment of heterogeneous group formation, was seen to be more moderately affected by the financial education programme.

On average, considering all types of students, no general effects of the different group formations and differentiated instruction on the course financial proficiency and financial knowledge measures were observed. However, for financial behaviour, it was found that homogeneous group

formation was more effective than heterogeneous group formation, yet, differentiated instruction did not improve performance. The results of the analysis established distinct heterogeneity of treatment effects, which is in line with the educational science literature on differentiation (e.g. Deunk et al., 2015; Lou et al., 1996; Slavin, 1990). Different patterns for students with different abilities and backgrounds were found. Suggestive evidence was obtained that the performance of low ability students improved most when a homogeneous group formation within the classroom was implemented and additional instructions were provided. Also, the scores of high ability students, who did not receive any additional instructions, still increased relatively more in case of differentiated instruction. Further, it was found that the non-native students significantly benefited from a homogeneous group formation within the classroom, given the material included additional instructions, with test scores increasing by 0.48 standard deviations compared to 0.22 and 0.37 standard deviations. On the other hand, the Dutch speaking students performed similar in all experimental conditions.

Two mechanisms are proposed through which this heterogeneous effect may be operating. That is, as non-native students are suggested to be less proficient in the teaching language, some of these students may have had more difficulty in following and understanding the lectures. Accordingly, differentiated instruction enabled to follow the programme at the student's own language-level (Figlio, Rush, & Lu, 2013). Alternatively, traditional standards might be set too high for language-minority students. Differentiated instruction mitigates the effects of these uniform standards (Alexander, 2000).

Students in large classrooms were also expected to benefit from differentiated instruction when compared to the students in smaller classrooms. Although it was observed that homogeneous group formation had a positive impact, adding differentiated instruction was not seen to increase the effectiveness, which is in contrast to the previous research (see for example Barrow, Markman, & Rouse, 2009). This finding, however, is consistent with several explanations. First, Kulik & Kulik (1982), demonstrated that group formations generally have no impact on student achievement, yet, a positive influence on students' attitude during class. A descriptive analysis of the classroom observations reveals this finding to be true in this paper as well, i.e. in classrooms where teachers implemented homogeneous group formation, students collaborated more with their peer and were more actively involved with the didactic material. Second, the lack of impact of differentiated instruction could also be attributed to the fact that while the programme was designed to deliberately minimise the role of teachers, yet, teachers might still have interacted with their students. If the teachers did not follow instructions as advised, say, they helped their students with the content of the material, then the effect of differentiated instruction implemented in the material itself was likely to be mitigated. Again, through descriptive analysis of the classroom observations, the occurrence of these events was reported for the majority of classroom observations. A limitation of this paper, however, is that this matter cannot be analysed empirically.[11] Finally, it is worth noting that due to the selective secondary education system in Flanders, students'

[11] 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.

abilities within classrooms were already relatively similar. Therefore, it may be suggested that the differentiated instruction implemented in the financial education programme is likely to be more effective in countries with a comprehensive school system having more heterogeneous classrooms.

The findings of this paper have important policy implications considering the low financial literacy levels and significant heterogeneity in the present day classrooms. First, it is shown that providing additional instructions to some students does not come at the expense of other students in the classroom when groups of students are formed homogeneously. As the performance of all the students with different ability levels improved, it may be concluded that high ability students were put under a more challenging situation as they received minimal instructions and had to find solutions in a self-regulated way. Second, bearing in mind the benefits of high financial literacy levels, the findings of this paper reinforce the need for the integration of financial literacy competencies into the education programme. Third, given that teachers are unable to meet the diverse needs of students while teaching *the average student*, the present paper provides causal evidence on the effectiveness of classroom differentiation practices for financial literacy tools. Manipulating the class composition and implementing basic adapted instructions were found to increase the learning outcomes of struggling students in classrooms without deteriorating the performance of their peers. These practices of differentiated teaching and learning deserve the attention of the policymakers as these may be more cost-effective and easier to implement than other educational interventions, such as tracking and class size reductions.

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.*

Alexander, N. A. (2000). The missing link: an econometric analysis on the impact of curriculum standards on student achievement. *Economics of Education Review*, *19*, 351–361.

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

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying Education : Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, 1235–1264.

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.

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.

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). CEM: Coarsened exact matching in Stata. *The Stata Journal*, *9*(4), 524-546.

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.

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*, 1739–1774.

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

Figlio, D., Rush, M., & Yin, L. (2013). Is It Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning. *Journal of Labor Economics*, *31*(4), 763–784.

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*.

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

Hanushek, E. A., & Wößmann, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries. *The Economic Journal*, *116*, 63–76.

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*.

Iacus, S., King, G., & Porro, G. (2008). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis, 20*(1), 1-24

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).

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment effects. *Review of Economic Studies*, *76*, 1071–1102.

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.

Lührmann, M., Serra-Garcia, M., & Winter, J. (2018). The Impact of Financial Education on Adolescents' Intertemporal Choices. *American Economic Journal: Economic Policy*, *10*(3), 309–332.

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.

Merlino, L. and Steinhardt, M. F. and Wren-Lewis, L. (2018). More than Just Friends? School Peers and Adult Interracial Relationships. *Journal of Labor Economics*, Forthcoming.

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

OECD. (2016). *PISA 2015 Financial Literacy Framework*. PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy, OECD Publishing, Paris.

Rouse, C. E., & Krueger, A. B. (2004). Putting computerized instruction to the test : a randomized evaluation of a 'scientifically based' reading program. *Economics of Education Review*, *23*, 323–338.

Schiltz, F., Mazrekaj, D., Horn, D., & De Witte, K. (2019). Does It Matter When Your Smartest Peers Leave Your Class? Evidence from Hungary. *Labour Economics*, In Press.

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

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.

van Rooij, M., Lusardi, A., & Alessie, R. J. M. (2011). *Financial literacy, retirement planning, and household wealth*. NBER Working Paper, No. 17339.

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.

Walstad, W., Urban, C., Asarta, C. J., Breitbach, E., Bosshardt, W., Heath, J., Neill, B., Wagner, J., & Jian, J. (2017). Perspectives on evaluation in financial education: Landscape, issues, and studies. *The Journal of Economic Education*, *48*(2), 93–112.

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.

Appendix

Appendix AI: Figures

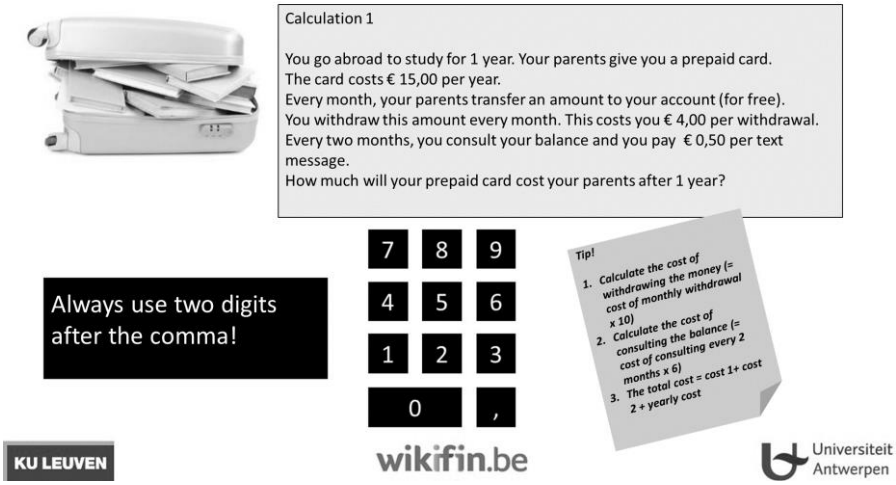


Figure AI: Differentiated instruction in didactic material

Note: Sticky note in didactic material indicates how to do the calculation needed in order to solve the problem.

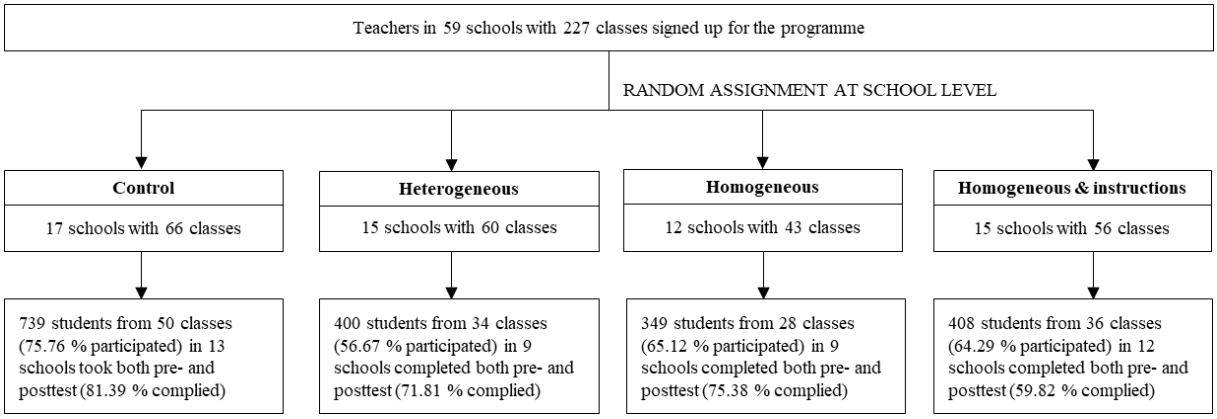


Figure AII: Selection of participants per treatment for Study I

Note: The final sample consists of 1,896 eighth grade students in 148 classes in 43 schools.

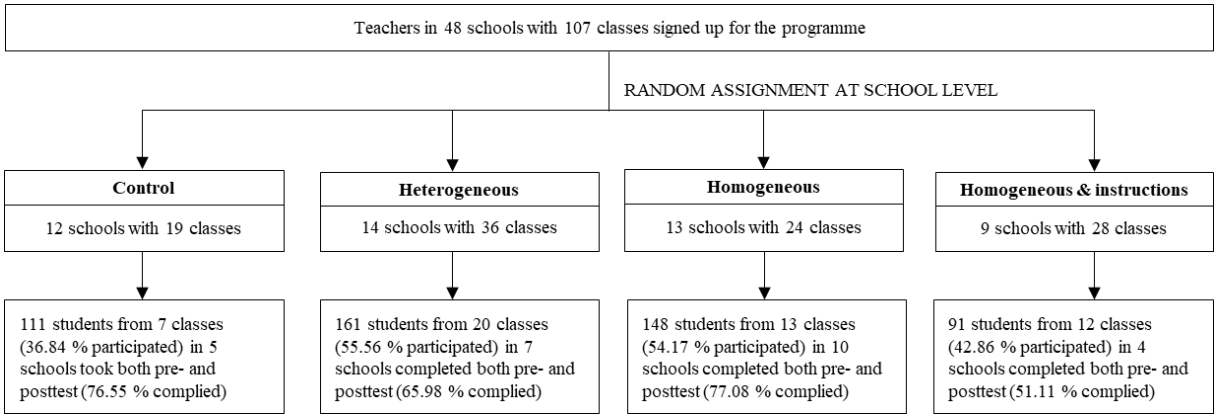


Figure AIII: Selection of participants per treatment for Study II

Note: The final sample consists of 511 ninth grade students in 52 classes in 26 schools.

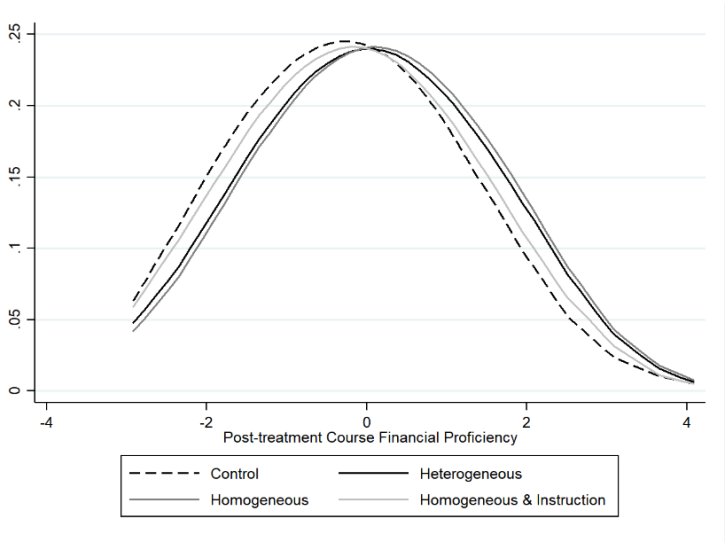


Figure AIV: Distribution Shift of Course Financial Proficiency

Note: Comparison of density plots for post-treatment course financial proficiency measure in experimental and control conditions. The Kolmogorov-Smirnov test rejects equality of distributions.

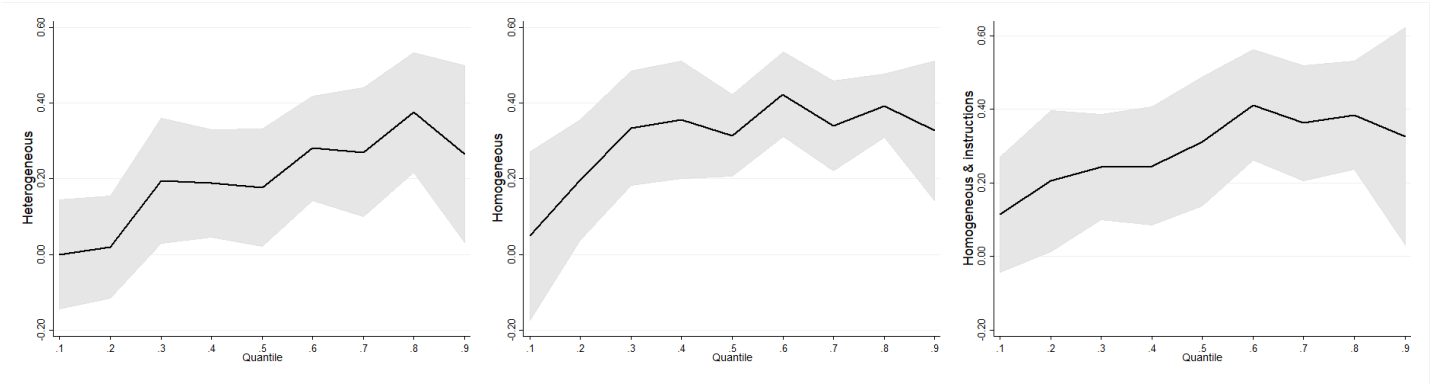


Figure AV: Quantile regressions for Course Financial Proficiency

Note: This figure plots the treatment coefficients of the quantile regression for the course financial proficiency measure and 95 percent confidence intervals. The estimated equation includes the following controls; gender, SES, track, type of school, and grade. The confidence intervals were obtained using bootstrap standard errors using 1000 bootstrap replications.

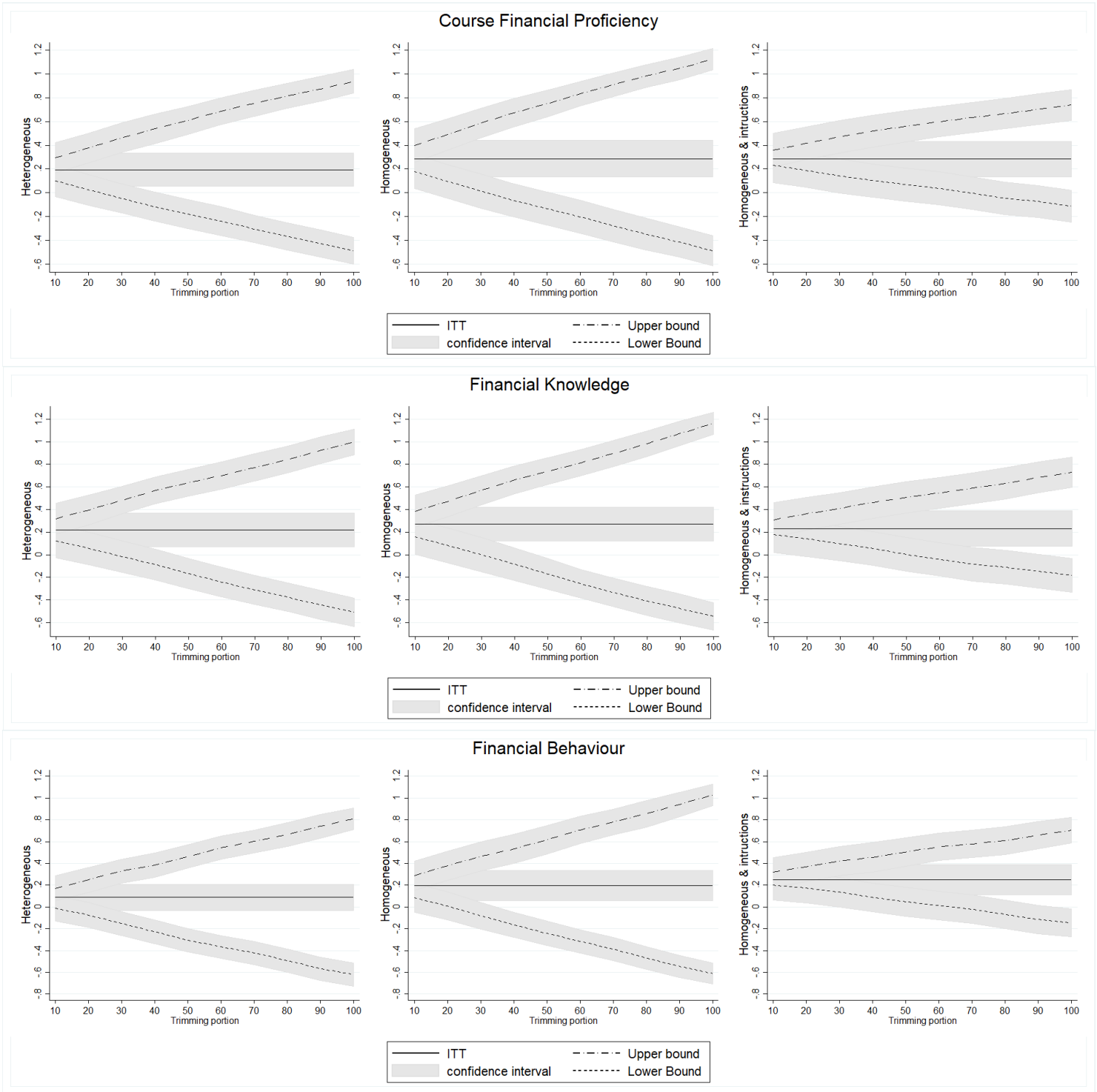


Figure AVI: Lee bounds for Study I & II

Note: The approach described in Merlino, Steinhardt, and Wren-Lewis (2018) is followed in order to compute the bounds for which the control condition and two experimental conditions that are least affected by selection (see figure AII and AIII) are manually trimmed. Particularly, for the computation of the upper (lower) bound, fractions of observations that contribute (most) least to a positive correlation between the dependent variable and treatment variables are dropped. That is, from the control condition, students with the highest (smallest) residuals when a regression was performed of the outcome value on baseline values and other demographics with standard errors clustered at class level are dropped. For the experimental conditions (Heterogeneous and Homogeneous), the samples are trimmed by dropping students with the smallest (highest) residuals. The bounds estimates are then derived from the re-estimation of ITT regression models on the trimmed sample. Further, a set of bounds are defined for different portions (from 10 to 100 percent) of the trimming fractions, corresponding to less versus more restrictive bounds. The graphs show the bounds and main ITT estimates with confidence intervals plotted against the different trimming portions for all three outcome measures.

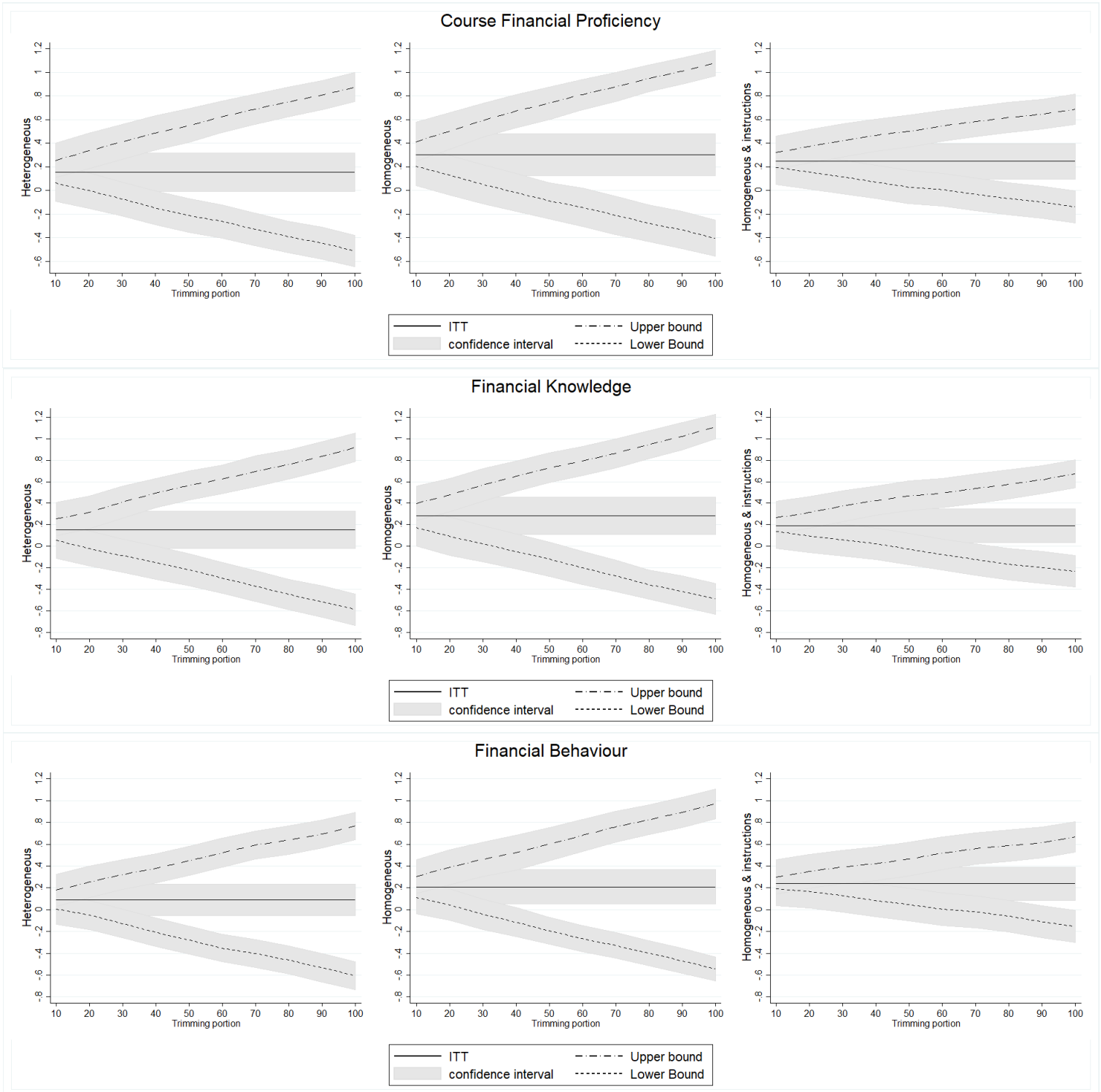


Figure AVII: Lee bounds for Study I

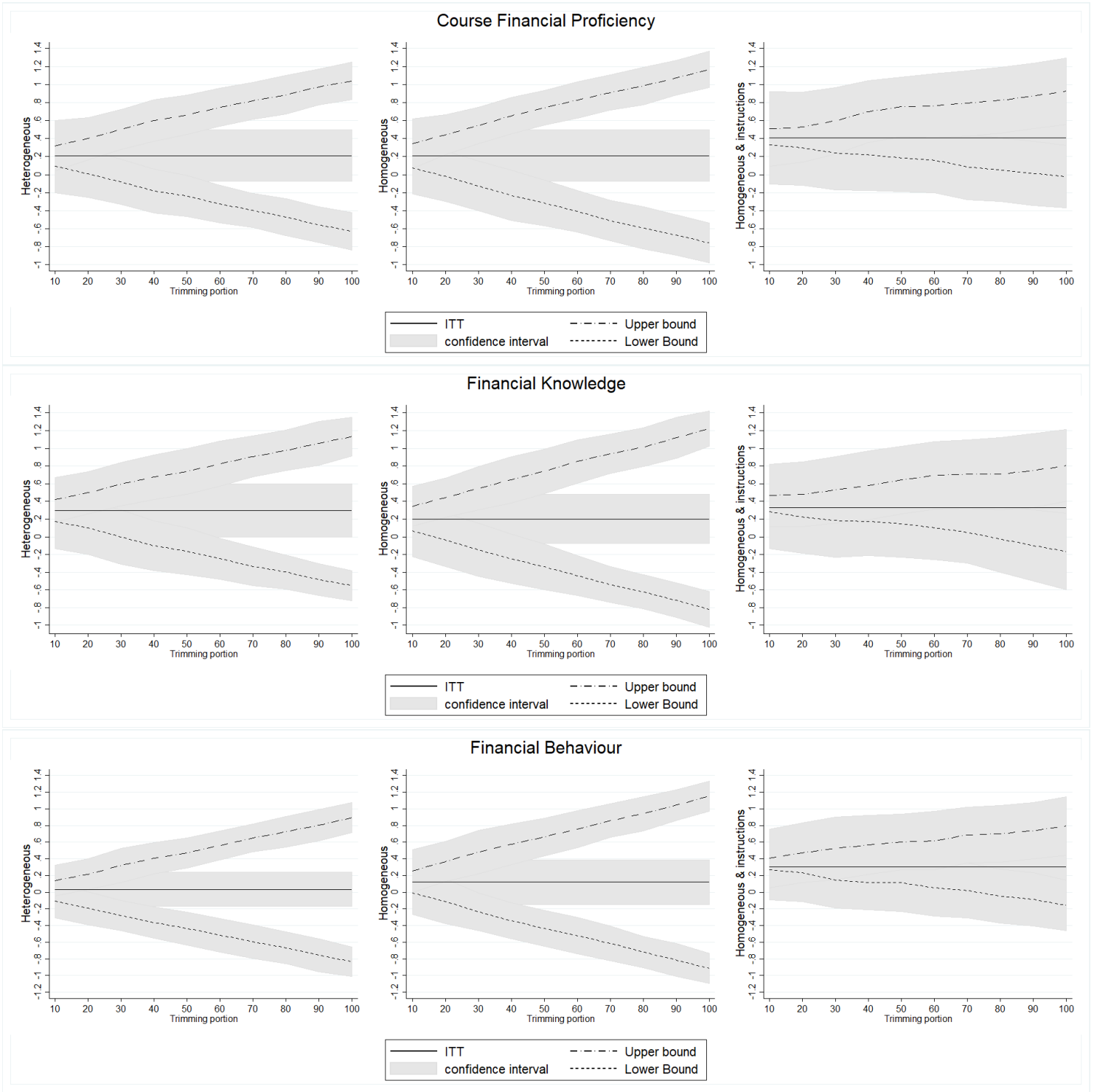


Figure AVIII: Lee bounds for Study II

Appendix AII: Tables

Table AI: Descriptive Statistics Study I & Study II

STUDY 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
Track	*Academic*	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								
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
Pre-treatment financial scores								
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
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
Post-treatment financial scores								
Course financial proficiency (9)		2.89 (1.66)	3.29 (1.75)	0.019	3.61 (1.71)	0.000	3.17 (1.69)	0.108
Financial knowledge (6)		1.91 (1.22)	2.21 (1.24)	0.009	2.38 (1.25)	0.000	2.10 (1.23)	0.091
Financial behaviour (3)		0.99 (0.89)	1.08 (0.93)	0.253	1.23 (0.94)	0.012	1.07 (0.93)	0.337
STUDY 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
Track	*Academic*	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								
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
Pre-treatment financial scores								
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
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

Post-treatment financial scores							
Course financial proficiency (9)	3.47 (1.48)	4.07 (1.73)	0.042	3.75 (1.74)	0.386	3.14 (1.84)	0.530
Financial knowledge (6)	2.19 (1.05)	2.66 (1.22)	0.033	2.39 (1.28)	0.385	2.04 (1.26)	0.672
Financial behaviour (3)	1.28 (0.85)	1.42 (0.93)	0.264	1.36 (0.93)	0.561	1.10 (0.97)	0.405

Note: Mean values and standard deviations in parentheses. For the track, absolute and relative frequencies are given. The p-values are derived from difference in means test between students in treatment and control schools (means of a regression of the characteristic on the treatment indicator with clustered standard errors at class level). Financial scores are reported before standardization.

Table AII: Baseline characteristics of Students

STUDY I	**Variables**	**Control**	**Heterogeneous**	*p*-value	**Homogeneous**	*p*-value	**Homogeneous & instruction**	*p*-value
Number of Schools		16	11		10		14	
Number of Students		908	557		463		682	
Type of school (private)		0.63	0.96	0.000	0.83	0.029	0.83	0.018
Class size		18.19 (4.90)	16.43 (5.35)	0.089	16.73 (5.45)	0.162	15.33 (5.33)	0.003
Track	*Academic*	774 (85.24%)	459 (82.41%)		396 (85.53%)		416 (61.00%)	
	Technical	47 (5.18%)	50 (8.98%)	0.862	23 (4.97%)	0.996	152 (22.29%)	0.014
	Vocational	87 (9.58%)	48 (8.62%)		44 (9.50%)		114 (16.72%)	
Background characteristics								
Gender (female)		0.52	0.56	0.261	0.45	0.264	0.50	0.759
Age (years)		14.19 (0.48)	14.17 (0.50)	0.766	14.17 (0.45)	0.673	14.31 (0.56)	0.016
Number of holidays per year (4)		2.90 (1.04)	3.03 (1.06)	0.105	2.90 (1.05)	0.980	2.85 (1.04)	0.485
Language (Dutch)		0.84	0.73	0.043	0.89	0.085	0.62	0.000
Ability		3.33 (1.03)	3.18 (1.03)	0.204	3.40 (0.89)	0.478	3.00 (1.08)	0.006
Financial characteristics & scores								
Importance of financial literacy (5)		4.14 (0.76)	4.15 (0.75)	0.875	4.22 (0.70)	0.069	4.04 (0.84)	0.089
Importance of saving (5)		4.21 (0.96)	4.15 (0.98)	0.372	4.35 (0.84)	0.020	4.05 (1.11)	0.017
Comparison shops before purchase (5)		3.60 (1.30)	3.54 (1.34)	0.473	3.77 (1.27)	0.024	3.60 (1.33)	0.928
Course financial proficiency (9)		2.90 (1.44)	2.86 (1.59)	0.741	3.19 (1.51)	0.039	2.88 (1.60)	0.848
Financial knowledge (6)		1.78 (1.07)	1.82 (1.14)	0.640	2.01 (1.15)	0.033	1.86 (1.17)	0.400
Financial behaviour (3)		1.12 (0.81)	1.03 (0.79)	0.189	1.18 (0.81)	0.347	1.02 (0.81)	0.114
STUDY II	**Variables**	**Control**	**Heterogeneous**	*p*-value	**Homogeneous**	*p*-value	**Homogeneous & instruction**	*p*-value
Number of Schools		7	10		9		6	
Number of Students		145	224		192		178	
Type of school (private)		0.93	0.89	0.646	0.90	0.732	0.73	0.147
Class size		16.41 (4.86)	12.01 (5.88)	0.042	16.76 (6.46)	0.887	10.93 (2.73)	0.003
Track	*Academic*	92 (63.45%)	191 (78.28%)		141 (73.44%)		48 (26.97%)	
	Technical	24 (16.55%)	39 (15.98%)	0.320	21 (10.94%)	0.670	79 (44.38%)	0.179
	Vocational	29 (20.00%)	14 (5.74%)		30 (15.63%)		51 (28.65%)	
Background characteristics								
Gender (female)		0.66	0.54	0.045	0.52	0.015	0.54	0.094
Age (years)		15.21 (0.51)	15.13 (0.46)	0.413	15.27 (0.54)	0.655	15.49 (0.64)	0.032
Number of holidays per year (4)		2.91 (1.05)	2.96 (1.05)	0.725	2.99 (0.98)	0.548	2.74 (1.11)	0.297
Language (Dutch)		0.82	0.86	0.522	0.85	0.670	0.66	0.058
Ability		3.12 (0.75)	3.22 (1.01)	0.513	3.27 (0.82)	0.325	3.17 (1.01)	0.733
Financial characteristics & scores								
Importance of financial literacy (5)		4.33 (0.65)	4.22 (0.71)	0.210	4.29 (0.68)	0.571	4.17 (0.74)	0.073
Importance of saving (5)		4.14 (1.02)	4.16 (0.96)	0.907	4.25 (0.92)	0.477	3.95 (1.10)	0.245
Comparison shops before purchase (5)		3.75 (1.30)	3.62 (1.27)	0.216	3.53 (1.26)	0.171	3.63 (1.31)	0.328
Course financial proficiency (9)		3.40 (1.69)	3.73 (1.52)	0.241	3.44 (1.51)	0.898	2.72 (1.76)	0.032
Financial knowledge (6)		2.14 (1.21)	2.37 (1.18)	0.256	2.14 (1.16)	0.984	1.64 (1.23)	0.018
Financial behaviour (3)		1.26 (0.82)	1.36 (0.85)	0.339	1.30 (0.86)	0.748	1.08 (0.87)	0.190

Note: Mean values and standard deviations in parentheses. For the track, absolute and relative frequencies are given. The p-values are derived from difference in means test between students in treatment and control schools (by means of a regression of the characteristic on the treatment indicator with clustered standard errors at class level). Financial scores are reported before standardization.

Table AIII: Comparison of schools

Characteristic	Participating schools (complying & non-complying)	Total population of schools	*p-value*
% low educated mothers	21,45	20.37	0.469
% on allowance	24,63	24.84	0.745
% language not Dutch	13,56	13.23	0.459
	Complying schools	**Total population of schools**	*p-value*
% low educated mothers	23.75	20.37	0.167
% on allowance	27.38	24.84	0.270
% language not Dutch	15.47	13.23	0.317
	Complying schools	**Non-complying schools**	*p-value*
% low educated mothers	23.75	23,58	0.091
% on allowance	27.38	27,69	0.079
% language not Dutch	15.47	14,20	0.365

Notes: Mean values and *p*-values of school composition characteristics when comparing different sets of schools derived from t-tests; *** p<0.01, ** p<0.05, * p<0.1; Participating schools refer to all schools which signed up for the programme, complying schools refer to schools where students completed both pre- and post-treatment tests.

Table AIV: Selective attrition

Dependent variable	Attrition Study I & II	Attrition Study I	Attrition Study II
Heterogeneous	0.160***	0.174***	0.118
	(0.0505)	(0.0534)	(0.123)
Homogeneous	0.0802	0.108	0.00577
	(0.0635)	(0.0699)	(0.147)
Homogeneous & instruction	0.219***	0.238***	0.156
	(0.0634)	(0.0672)	(0.163)
Controls	Yes	Yes	Yes
Obs.	3,369	2,610	759

Note: Standard errors clustered at class level in parentheses; *** p<0.01, ** p< 0.05, * p<0.1, Controls: pre-treatment value of outcome, gender, SES, track, type of school, (grade when Study I & II combined); Significant differences across treatment coefficients calculated by means of F-tests: Homogeneous and Homogeneous & instructions for study I & II ($p = 0.05$).

Table AV: Comparison of Non-Complying and Complying Students Within Conditions

STUDY I	**Variables**	**Control**	*p*-value	**Heterogeneous**	*p*-value	**Homogeneous**	*p*-value	**Homogeneous & instruction**	*p*-value
Percentage of Attrition		14.61%		28.19%		24.62%		40.18%	
Number of Attrited Students		169		157		114		274	
Type of school (private)		-0.31 (0.13)	0.020	-0.09 (0.06)	0.138	-0.37 (0.14)	0.015	-0.15 (0.09)	0.110
Class size		-2.27 (1.23)	0.071	-3.33 (1.04)	0.003	-2.66 (1.94)	0.177	-3.48 (0.98)	0.001
Track	*Academic*	-16.76%		-26.94%		-8.73%		-7.40%	
	Technical	12.54%	0.234	3.46%	0.009	3.88%	0.314	-2.48%	0.295
	Vocational	4.22%		23.40%		4.84%		9.88%	
Background characteristics									
Gender (female)		-0.11 (0.08)	0.179	-0.10 (0.06)	0.106	-0.01 (0.09)	0.872	-0.002 (0.05)	0.964
Age (years)		0.17 (0.06)	0.008	0.22 (0.09)	0.021	0.20 (0.10)	0.055	0.13 (0.08)	0.146
Number of holidays per year (4)		-0.10 (0.10)	0.328	-0.32 (0.14)	0.025	0.02 (0.13)	0.856	-0.16 (0.12)	0.206
Language (Dutch)		-0.03 (0.04)	0.514	-0.05 (0.07)	0.474	-0.06 (0.06)	0.262	-0.05 (0.07)	0.498
Ability		0.04 (0.12)	0.760	-0.04 (0.12)	0.739	-0.09 (0.15)	0.547	-0.30 (0.14)	0.042
Financial characteristics & scores									
Importance of financial literacy (5)		-0.09 (0.06)	0.163	-0.11 (0.08)	0.185	-0.13 (0.08)	0.113	-0.03 (0.09)	0.745
Importance of saving (5)		-0.11 (0.10)	0.310	-0.17 (0.11)	0.119	0.02 (0.10)	0.798	-0.16 (0.10)	0.123
Comparison shops before purchase (5)		0.05 (0.10)	0.609	-0.08 (0.15)	0.599	0.02 (0.16)	0.921	-0.19 (0.11)	0.095
Course financial proficiency (9)		-0.05 (0.16)	0.763	-0.54 (0.23)	0.022	-0.01 (0.26)	0.958	-0.20 (0.19)	0.306
Financial knowledge (6)		0.0003 (0.11)	0.998	-0.31 (0.14)	0.035	-0.08 (0.19)	0.658	-0.18 (0.13)	0.189
Financial behaviour (3)		-0.05 (0.09)	0.593	-0.22 (0.10)	0.038	0.07 (0.12)	0.567	-0.02 (0.09)	0.813

STUDY II	**Variables**	**Control**	*p*-value	**Heterogeneous**	*p*-value	**Homogeneous**	*p*-value	**Homogeneous & instruction**	*p*-value
Percentage of attrition		23.45%		34.02%		22.92%		48.89%	
Number of Attrited Students		34		83		44		87	
Type of school (private)		0.05 (0.06)	0.421	-0.18 (0.13)	0.179	-0.10 (0.12)	0.434	-0.19 (0.19)	0.333
Class size		-5.26 (2.31)	0.049	-0.27 (2.32)	0.909	1.79 (3.84)	0.650	-0.05 (1.02)	0.963
Track	*Academic*	-67.52%		-42.06%		7.93%		21.45%	
	Technical	36.01%	0.041	37.85%	0.047	-8.29%	0.810	-37.35%	0.877
	Vocational	31.51%		4.08%		0.37%		15.9%	
Background characteristics									
Gender (female)		0.10 (0.13)	0.489	-0.16 (0.08)	0.043	0.01 (0.07)	0.894	-0.03 (0.09)	0.719
Age (years)		0.27 (0.24)	0.297	0.17 (0.09)	0.061	0.01 (0.19)	0.963	0.12 (0.18)	0.495
Number of holidays per year (4)		-0.50 (0.17)	0.019	-0.25 (0.16)	0.141	0.12 (0.09)	0.176	0.18 (0.25)	0.485
Language (Dutch)		0.04 (0.09)	0.642	-0.11 (0.05)	0.041	0.01 (0.08)	0.879	-0.07 (0.10)	0.495
Ability		0.14 (0.21)	0.543	-0.18 (0.20)	0.381	0.40 (0.30)	0.213	-0.33 (0.19)	0.096
Financial characteristics & scores									
Importance of financial literacy (5)		-0.05 (0.19)	0.801	-0.24 (0.10)	0.027	-0.02 (0.09)	0.848	0.16 (0.11)	0.136
Importance of saving (5)		0.32 (0.22)	0.179	-0.14 (0.16)	0.366	0.09 (0.16)	0.592	-0.06 (0.17)	0.740
Comparison shops before purchase (5)		0.55 (0.15)	0.005	-0.19 (0.16)	0.239	0.46 (0.18)	0.026	0.46 (0.20)	0.035
Course financial proficiency (9)		-0.64 (0.35)	0.099	-0.75 (0.33)	0.030	-0.13 (0.26)	0.588	0.32 (0.36)	0.382
Financial knowledge (6)		-0.38 (0.24)	0.153	-0.32 (0.21)	0.141	-0.24 (0.20)	0.247	0.28 (0.22)	0.220
Financial behaviour (3)		-0.26 (0.15)	0.128	-0.42 (0.15)	0.010	0.12 (0.10)	0.273	0.05 (0.19)	0.802

Note: Differences in means between attrited and non-attrited students (completing the post-test) are obtained from regressing each characteristic on an indicator of attrition for which standard errors are clustered at class level (in parentheses). For the track, relative differences are given.

Table AVI: Financial attitudes of ninth grade students

Dependent variable	Financial attitudes					
	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	-0.74***	-0.579*	-0.404	-0.416	-0.57***	-0.365
	(0.288)	(0.312)	(0.317)	(0.349)	(0.193)	(0.288)
Controls	No	Yes	No	Yes	No	Yes
Obs.	504	504	504	504	504	504

Note: Estimates are derived from ordered logistic regressions; Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Controls: pre-treatment value of outcome, gender, SES, track, type of school.

Table AVII: Second post-treatment test

Dependent variable	Course Financial Proficiency		Financial Knowledge		Financial Behaviour	
Heterogeneous	0.180	0.196	0.0561	0.0965	0.253	0.231
	(0.154)	(0.179)	(0.127)	(0.145)	(0.155)	(0.169)
Homogeneous	-0.0163	-0.0480	-0.0965	-0.102	0.166	0.0992
	(0.118)	(0.118)	(0.0968)	(0.104)	(0.131)	(0.114)
Homogeneous & instructions	0.0532	0.273*	0.0388	0.264*	0.0565	0.182*
	(0.150)	(0.146)	(0.147)	(0.153)	(0.113)	(0.100)
Controls	No	Yes	No	Yes	No	Yes
Obs.	1,208	1,208	1,208	1,208	1,208	1,208

Note: Standard errors clustered at class level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Controls: pre-treatment value of outcome, gender, SES, track, type of school, grade, time between post-treatment tests; Significant differences across treatment coefficients calculated by means of F-tests: Homogeneous and Homogeneous & instructions for course financial proficiency ($p = 0.07$) and financial knowledge measure ($p = 0.046$).

Table AVIII: Robustness test – School Level Clustered Standard Errors

Dependent variable	Course Financial Proficiency		Financial Knowledge		Financial Behaviour	
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	No	Yes	No	Yes	No	Yes
Obs.	2,407	2,407	2,407	2,407	2,407	2,407

Note: Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Variables controlled for: pre-treatment value of outcome, gender, SES, track, type of school, grade; Significant differences across treatment coefficients calculated by means of F-tests - Heterogeneous and Homogeneous ($p = 0.037$) and Heterogeneous and Homogeneous & instructions ($p = 0.051$) for financial behaviour measure.

Table AIX: Robustness test – Instrumental Variables

Dependent variable	Course Financial Proficiency		Financial Knowledge		Financial Behaviour	
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	No	Yes	No	Yes	No	Yes
Obs.	1,516	1,516	1,516	1,516	1,516	1,516

Note: Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Variables controlled for: pre-treatment value of outcome, gender, SES, track, type of school, grade.

Table AX: Robustness test – Coarsened Exact Matching

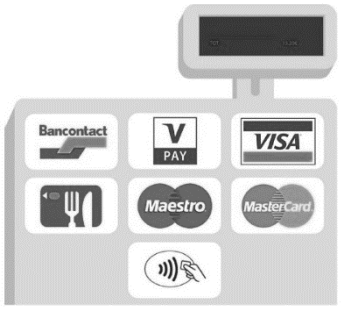
Dependent variable	Course Financial Proficiency		Financial Knowledge		Financial Behaviour	
	ITT	***CEM***	***ITT***	***CEM***	***ITT***	***CEM***
Heterogeneous	0.166**	0.204**	0.182**	0.288***	0.0788	0.121*
	(0.0715)	(0.101)	(0.0780)	(0.0883)	(0.0608)	(0.0688)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1,411	860	1,411	927	1,411	978
Homogeneous	0.254***	0.314***	0.251***	0.313***	0.159**	0.217**
	(0.0744)	(0.0948)	(0.0758)	(0.0931)	(0.0666)	(0.0903)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1,347	908	1,347	944	1,347	1,036
Homogeneous & instructions	0.297***	0.368***	0.233***	0.270***	0.260***	0.224**
	(0.0783)	(0.101)	(0.0845)	(0.0925)	(0.0707)	(0.0954)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1,349	734	1,349	780	1,349	905

Note: Estimates are derived from three seperate regressions including binary treatment variables; Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Variables controlled for ITT: pre-treatment value of outcome, gender, SES, track, type of school, grade; Variables controlled for CEM: baseline outcome values, class size, importance of financial literacy (for the middle and third panel). Matching was done based on the type of school, class size, track, gender, SES, pre-treatment value of outcome, importance of financial literacy, and grade.

Appendix AIII: Test instruments

First post-treatment financial proficiency test

1. *You find the following sticker in a department store. Indicate all correct claims.*



   - You can pay with eco vouchers
   - You can pay with bitcoins
   - You can pay with a cheque
   - You can pay with a credit card
   - You can pay contactless
   - You can pay with a debit card

2. *If you pay with a credit card, the money comes directly out of your checking account.*
   - Correct
   - False
   - I don't know
3. *Suppose that your parents have instructed their bank to automatically transfer the monthly rent from their checking account to the checking account of the landlord. Such a payment is a:*
   - Standing order
   - Bank transfer
   - Direct debit
   - I don't know
4. *Rank the following sources from most to least neutral (i.e. non-commercial).*
   a. [website of the government]
   b. [website of a bank]
   c. [website of a private consumer organisation or price comparison website]
   - a b c
   - a c b
   - b a c
   - c b a
   - I don't know
5. *Ann buys a book for €18. At checkout, she receives a discount of 10%. She pays with a bank note of €20. How much change will she get back?*
   - €3.80
   - €1.80
   - €16.20
   - €2.00
   - I don't know
6. *Marc does not own a checking account at IMG. He receives the following e-mail:*

From: IMG <imghelpdesk4@t-online.de>
To:
Date: Mon, 20 Nov 2017 23:02:33 +0100
Subject: Request new card

Dear client,

We have noticed from our administration that you are currently not using our new debit card. The new debit card is better secured against fraudulent activities and meets EU safety requirements. The requirements state that banks must now secure their clients against financial loss in case of potential fraudulent activities. Therefore, we would like to introduce you this more secured debit card.
Due to safety reasons, IMG obliges all her clients to use the new debit card.

Click here to order your new debit card.

We trust to have informed you sufficiently.

Thank you in advance for your cooperation,
Kinds regards,
Jeroen Goossens
IMG client service

Indicate which claim(s) are good advice for Marc.

- Open a checking account at this bank such that you can order the free debit card
- Ignore the e-mail, delete it from the mailbox and potentially warn the bank about this e-mail
- Click on the link to order the debit card
- Reply to this e-mail asking for more information
- Contact Jeroen Goossens asking for more information concerning this e-mail
- I don't know

7. *When you find the following logo (Se-cure) at the cash register, it means that:*

- You can pay with meal vouchers
- You can pay with your smartphone
- The store provides safe payment systems
- I don't know

8. *What does 'being careful with your bank card' involve?*
   - Entering your credit card details on a secured website of a known and reliable dealer
   - Writing your PIN on the back of your debit card
   - Confirming an online payment preferably with a code you insert on your card reader
   - Putting your bank card together with your PIN next to the card reader
   - I don't know
9. *Indicate which of the following persons take care with the different methods of payment.*
   - Mia can safely give her debit card and PIN to her friends
   - Paul puts his debit card, which he pays contactless with, in a metal case
   - Elias immediately calls Card Stop when he loses his debit card
   - Helene does not log out in the event she makes a payment with her smartphone because the app closes automatically
   - I don't know