

**The Effect of Ability Matching
and Differentiated Instruction
in Financial Literacy Educa-
tion. Evidence from Two
Randomised Control Trials.**

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

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

www.cesifo-group.org/wp

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

The Effect of Ability Matching and Differentiated Instruction in Financial Literacy Education. Evidence from Two Randomised Control Trials.

Abstract

The present paper examines the impact of ability matching 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 ability matching is investigated by forming pairs of students either randomly or based on their abilities. In addition, the paper studies whether differentiated instruction, in the form of additional instructions for lower ability students, enhances the impact of the financial education programme. The paper provides evidence on the effects of both practices using two randomised control trials involving 65 schools and 2,407 students. Overall, the results suggest that the programme raises the financial proficiency of students by 0.18 standard deviations. Although the learning outcomes of the average student are not affected by the differentiation practices, non-native students significantly benefit from differentiated instruction.

JEL-Codes: C930, I210, O160.

Keywords: financial literacy, ability matching, differentiated instruction, randomised control trial.

Kaat Iterbeke
Leuven Economics of Education Research
KU Leuven / Belgium
kaat.iterbeke@kuleuven.be

Koen Declercq
Leuven Economics of Education Research
KU Leuven / Belgium
koen.declercq@kuleuven.be

Kristof De Witte
Leuven Economics of Education Research
KU Leuven / Belgium
kristof.dewitte@kuleuven.be

Wouter Schelfhout
Antwerp School of Education
Universiteit Antwerpen / Belgium
wouter.schelfhout@uantwerpen.be

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, Lieve Lammens, Johan Mestdagh, and Danièle Van der Espt for their research assistance. We thank conference and seminar participants at LEER 2018, LEER 2019, LAGV 2019, ESPE 2019, EEA 2019, EALE 2019, and Politecnico di Milano for valuable suggestions on an earlier version of this paper.

1. Introduction

Financial literacy levels are low among the young and persist over the life cycle (Lusardi & Mitchell, 2014). Previous literature reveals negative impacts on household's decision making when financial knowledge is poor. That is, individuals with low financial literacy levels are less likely to engage in the stock market and plan for retirement, and hence, less likely to accumulate and manage wealth effectively (van Rooij, Lusardi, & Alessie, 2012). Consequently, in face of today's increasingly complex financial landscape and the detrimental effects of low financial literacy levels, governments, non-profit organisations, and financial institutions have started to promote the development of financial education programmes. Although former remedial approaches targeted merely adults, the focus has shifted to school-based interventions. Apart from financial illiteracy being more pronounced among youth (e.g. Lusardi, Mitchell, & Curto, 2010; Mandell, 2008), programmes offered at school are attractive for several reasons. Financial guidance can be provided when students are still developing habits and have not made major financial decisions as adults. Furthermore, selective participation is minimised and spillover effects can be generated as students are likely to act as an agent of change in their family's financial decisions (Frisancho, 2018; Walstad et al., 2017). Along with the policy attention, a large body of academic research has emerged discussing the impact of financial education programmes among youth. Recent experimental studies established that financial education programmes increase financial knowledge and also affect financial attitudes and behaviour of students.¹

Nevertheless, despite the initiatives dealing with the inadequate financial literacy levels among youth, large heterogeneity in financial literacy is observed. Along with sociodemographic, family and peer characteristics, cognitive ability is found to play a role in explaining the differences in financial knowledge (OECD, 2017; Lusardi, Mitchell, & Curto, 2010). Given the differences prevailing in classrooms, a differentiated approach, aimed at addressing the particular needs of *all* students, may be more effective. Using a randomised evaluation, the present paper examines the effects of two within-class differentiation practices on student performance, i.e. ability matching in student pairs and differentiated instruction (in the form of additional instructions for lower ability pairs). To the extent that students benefit from high-achieving peers, grouping students into ability-matched pairs may disadvantage low-ability students, thereby exacerbating inequalities (Epple, Newlon, & Romano, 2002). Yet, ability matching allows to appropriately address students' variance and match instruction to students' needs, benefiting *all* students (Duflo, Dupas, & Kremer, 2011). Consequently, as suggested in the educational science literature (Lou et al., 1996; Slavin, 1990), we hypothesize that struggling students benefit from ability-matched pairing only if instructions are adapted to their specific needs, whereas high achievers perform well in ability-matched pairs regardless of the level of instructions.

¹ Such as Becchetti, Caiazza, & Coviello, 2013; Becchetti & Pisani, 2012; Berry, Karlan, & Pradhan, 2018; Bruhn, de Souza Leão, Legovini, Marchetti, & Zia, 2016; Frisancho, 2018; Hinojosa, Miller, Swanlund, Hallberg, Brown, & O'Brien, 2010; Lührmann, Serra-Garcia, & Winter, 2015, 2018; Villanueva, Bover, & Hospido, 2018.

Our paper contributes to the literature by evaluating the causal impact of ability matching and differentiated instruction on student learning outcomes. In particular, we contribute 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. To the best of our knowledge, our research is novel in its approach since previous evaluations of financial education programmes have not considered the heterogeneity of students' performances in financial literacy in the programme design itself, i.e. all students were being taught the same programme in a uniform way. From a broader policy perspective, examining the effectiveness of within-class differentiation practices 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). Furthermore, as tracking is particularly harmful to students from disadvantaged backgrounds (Hanushek & Woessman, 2006), differentiation might avoid the negative impact on these subgroups.

To test the research question, two randomised control trials were set up in the eighth and ninth grade of secondary school in Flanders, the northern region of Belgium. Examining the different age groups not only enables us to observe age-specific heterogeneity, the validity of results can be examined through it as well, 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 an interactive learning game for which students were made to work in pairs. By means of a random assignment of schools to specific conditions, we evaluate the effect of the programme and study whether ability matching and differentiated instruction further enhance the learning outcomes. 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 ability matching, the students were matched according to their abilities in the second experimental condition. Finally, to investigate the effectiveness of differentiated instruction, we considered ability matching with additional instructions to lower ability pairs in the third experimental condition.

The findings of our paper suggest that the financial education programme is effective, i.e. the programme raises the financial proficiency of students by 0.18 standard deviations.² Although no major effects of ability matching and differentiated instruction are found for the average student, strong heterogeneous effects are observed as non-native students benefit from being matched to a peer with similar ability, conditional on receiving differentiated instruction.

² Financial proficiency was measured via a computer-aided multiple-choice test consisting of nine questions on both financial knowledge and behaviour that referred directly to the content taught during the programme.

The remaining paper unfolds as follows. The next section describes the design including the institutional context, the procedure, test instruments, and the empirical strategy. Section 3 discusses the data. Section 4 and 5 present the main results and test for the robustness of results, respectively. Section 6 concludes.

2. Experimental design

We conducted two randomised control trials involving secondary school students in eighth and ninth grade in the region of Flanders. The final sample includes 2,407 students in 200 classes in 65 schools. The first experimental study took place between January, 2018 and May, 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 the ninth grade and covers data on 511 students in 52 classes in 26 schools.³ The second study started in February, 2018 and ended in June, 2018.

In this section, we first discuss the institutional context and the experimental procedure including an overview of the study timeline, the content and implementation of the financial education programme, and details on the differentiation practices and experimental conditions. Further, the test instruments and empirical strategy are presented.

2.1. Institutional Context

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. Further, Flemish secondary education is organized in a tracking system. Students can choose between programmes in an academic, technical, artistic, or vocational education track.⁴

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 in the content of the material taken for this paper.⁵

³ A small number of schools participated in both studies.

⁴ Officially, the distinction between tracks exists only from the ninth grade onwards. However, at the start of the eighth grade, most schools allow students to decide on elective courses that prepare for a particular education track.

⁵ It should be noted that financial literacy competencies became part of the education curriculum in Flanders from September 2019 onwards (De Witte, Holz & De Becker, 2019).

2.2. 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. Randomisation at school level has advantages and disadvantages. On the one hand, in the design of the studies, we chose for randomisation at school level as it ensures minimal interference for schools participating with multiple classes. In other words, it allows us to avoid contamination effects, which are very difficult to measure and control for, and which would create serious endogeneity issues. On the other hand, randomisation at school level might result in differences at individual level across conditions as students within the same school are more homogeneous compared to students in other schools.

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 from parents, the material had to be delivered during regular class hours. Second, to measure the impact of the programme, all students had to take three financial literacy tests, i.e. a pre-treatment test prior to the financial education programme, measuring the baseline financial proficiency of students, and two post-treatment tests, capturing potential short- and long-term impacts of the programme. 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 assignment approximately six weeks after the lectures. The homework assignment facilitates the collection of data, but comes at the cost of losing control over whether the results reflect actual knowledge and behaviour. Although the instructions explicitly mentioned that students should take the test on their own, there might have been collaboration between the student and other individuals (e.g. peers or parents). However, as long as these contaminations are random across the conditions, the test will remain valid. 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	Mid-End Jan	End Jan		End Mar	May
<i>Treatment schools</i>	Randomisation to an experimental condition	Student pre-test	Material sent to schools	<i>6/8 weeks to teach 4 lectures</i>	1 st post-test & teacher survey	2 nd post-test
<i>Control schools</i>	Randomisation to control condition	Student pre-test			1 st post-test & teacher survey	
STUDY II	End Feb	End Feb-Mar	Mid Apr		Jun	End Jun

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 ability matching and differentiated instruction in financial literacy education, schools were randomly allocated to four different conditions;

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 was designed by senior teachers in the research team as a computerised serious game consisting of four lectures of 50 minutes each on the topic ‘means of payment’. This format was used in order to create an interactive environment that 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.⁶ 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 the 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 that 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 classes, the role of the teacher was reduced, such that potential additional teacher effects on the effectiveness of the financial education programme were minimised by design.⁷

Matching. The literature suggests two forms of matching to handle differences in classrooms (Deunk et al., 2015). On the one hand, for randomly-matched pairs, students were assigned a peer through a matching game at the start of the course in which, after the teacher had laid out a set of pairs of images face down in a random order, each student had to pick a card and was matched with the student who chose the same image. On the other hand, for ability-matched pairs, the students were differentiated based on their ability. In particular, before the start of the course, teachers were requested to form pairs

⁶ 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.

⁷ 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-based version in their classroom. Results from a regression analysis indicate no difference in the effectiveness of the computerised or paper-based version of the didactic material.

of students according to their performance in mathematics in the previous semester.⁸ 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 explains more than 70 percent of the variation in financial literacy performance (OECD, 2017). Hence, we expected that the high ability pairs for which ability-matching was implemented would perform better during the lectures.

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 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. This enabled us to coach their learning process adaptively (see Figure A.I in Appendix). Note that, in practice, if differentiated instruction was implemented in the classroom, the teacher was asked to split the student pairs into three equal groups according to their math ability. Finally, it should be noted that this form of differentiated instruction may either result in convergence or divergence of performances in a classroom, i.e. 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.⁹ Hence, the heterogeneity of students' test scores will be examined in the subsequent analysis in order to determine whether differentiated instruction resulted in convergence or divergence of the test scores.

Experimental conditions. Although the financial education programme was similar for all 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. In the first experimental condition, a random selection of schools was given the material that involved randomly-matched student pairs and no differentiated instruction. In the second experimental condition, the same financial education programme was used, but the students were matched according to their ability and no differentiated instruction was provided to these pairs. Comparing students in these two experimental conditions allows us to identify the impact of ability matching on learning outcomes. In the third condition, ability matching was used along with

⁸ It should be noted that teachers may have incorporated additional factors in forming these groups given they may know the students and know who may or may not work well together. Although this would contaminate the 'ability match' treatment, teachers were asked to follow a strict protocol (i.e. 'Students do not have the liberty to choose themselves, but are assigned a peer based on their grade in mathematics of the previous semester. In practice, you proceed as follows. Match students according to increasing grades in mathematics: first the two weakest, then the next two weaker students, and so on until the strongest two students of the class are paired.'). Classroom observations suggest a correct implementation.

⁹ Teachers aiming at convergence wish all students in the classroom to reach a minimum level of performance, whereas divergence refers to helping all children reach their highest potential (Deunk et al., 2015).

differentiated instruction to evaluate whether differentiated instruction enhances the impact of the programme for ability-matched student pairs. Table II presents a schematic overview of the different experimental conditions.

Table II: Control and experimental conditions

	Control condition	Experimental condition I Random match	Experimental condition II Ability match	Experimental condition III Ability match & instruction
Matching	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.

2.3. Test instruments

We use a pre-treatment test and two post-treatment tests to assess the impact of the financial education programme on student financial proficiency. The computer-aided multiple-choice tests consisted of nine questions that related directly to the material on means of payment. The pre-treatment financial proficiency test also consisted of a short survey on demographics and statements related to financial attitudes. The post-treatment tests consisted of nine similar questions measuring the same concepts with slight rephrasing and number adjustments (see Appendix C for a detailed description). For the pre- and first post-treatment test, the duration was eleven minutes on 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’, financial knowledge, and financial behaviour. The ‘course financial proficiency’ measure combines the financial knowledge and behaviour measures. Students’ financial knowledge is measured using 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 C). Financial behaviour is assessed using three questions related to students’ views on being careful with different means of payment and information (questions 6, 8, and 9 in Appendix C).

The behaviour of students and teachers is assessed by classroom observations. After emails were sent to schools for consent, three experienced PhD researchers examined 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 lectures were filmed. The classroom observations provide evidence on the exact implementation in the schools.

2.4. Empirical strategy

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:

$$y_{ijs}^1 = \alpha + \beta_0 \text{Treat}_{is} + \beta_1 \text{Treat}_{is} * \text{matching}_{is} + \beta_2 \text{Treat}_{is} * \text{instruction}_{is} + \beta_3 y_{ijs}^0 + \sum \beta_4' S_s + \sum \beta_5' X_i + \epsilon_{ij} \quad (1)$$

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 ; $Treat_{is}$ represents a dichotomous variable indicating whether student i in school s is assigned to the control or one of the three experimental conditions and corresponds to the effect of random matching; the interaction term $Treat_{is} * matching_{is}$ measures the effect of ability matching as compared to random matching for student i in school s ; the interaction term $Treat_{is} * instruction_{is}$ indicates whether providing additional differentiated instruction to ability-matched pairs enhances the effect of the programme for student i in school s ; 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 public or private school) for school s ; X_i indicates individual characteristics (such as the gender, socioeconomic status, and education track) of student i and is included to aid precision and control for potential baseline imbalances; ϵ_{ij} is a clustered error term at class level j for student i . Clustering is 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 findings (see Table B.VI). All test measures are standardized and hence, regression coefficients can be interpreted in terms of standard deviation changes.

3. Data

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

3.1. 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 B.I). In total, 2,407 students in 65 schools were randomly allocated to one of the four conditions. Each condition contains about 16 schools. The third, fifth, and seventh columns in Table III present the p -values of a regression of the characteristic on the treatment indicator (using clustered standard errors at class level) and signals the differences in school and student characteristics between the control and each experimental condition. Overall, the students in treatment and control schools appear to be relatively similar, however, some significant differences can be observed. 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 study track, yet this share is the lowest in the experimental condition of ability matching with differentiated instruction.

The second panel of Table III shows the background characteristics of students. The sample contains slightly more girls than boys and the students are, on average, 14 years old. The proxy for the socioeconomic status (SES) 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.¹⁰ Most students speak Dutch at home (the official language in Flanders), however, in the third experimental condition this share is significantly lower. Finally, the variable ‘grade in mathematics’ is measured on a five-point scale and merely provides a proxy for the performance of students in mathematics as 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	Random match	<i>p</i> - value	Ability match	<i>p</i> - value	Ability match & instruction	<i>p</i> - value
<i>Number of Schools</i>	17	16		17		15	
<i>Number of Students</i>	850	561		497		499	
Type of school (private)	0.72	0.97	0.000	0.92	0.007	0.88	0.036
Class size	18.48 (4.63)	15.86 (5.71)	0.008	17.08 (5.14)	0.146	15.68 (5.32)	0.005
Track							
<i>Academic</i>	741 (87.2%)	509 (90.73%)		412 (82.90%)		276 (55.31%)	
<i>Technical</i>	35 (4.12%)	39 (6.95%)	0.214	37 (7.44%)	0.646	117 (22.85%)	0.002
<i>Vocational</i>	74 (8.71%)	13 (2.32%)		48 (9.66%)		109 (21.84%)	
<i>Background characteristics</i>							
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
Grade in mathematics (5)	3.06 (1.29)	3.05 (1.26)	0.911	3.23 (1.13)	0.150	2.98 (1.31)	0.583
<i>Financial characteristics</i>							
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)	3.60 (1.32)	0.983	3.66 (1.28)	0.407	3.63 (1.33)	0.700
<i>Pre-treatment financial scores</i>							
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
<i>1st post-treatment financial scores</i>							
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
<i>2nd post-treatment financial scores</i>							
Course financial proficiency (9)		3.25 (1.54)		3.12 (1.66)		3.06 (1.65)	
Financial knowledge (6)		1.98 (1.15)		1.92 (1.19)		2.02 (1.24)	
Financial behaviour (3)		1.27 (0.73)		1.20 (0.79)		1.05 (0.78)	

Note: Mean values and standard deviation in parentheses; For track, absolute and relative frequencies are given; *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 standardisation; Students from control schools did not complete the second post-treatment test. Accordingly, their first post-treatment test scores will be used as a comparison in the analysis of the second post-test, considering no learning effects of the test itself.

¹⁰ Maldonado, De Witte and Declercq (2019) show the validity of this proxy as it correlates strongly to the number of disadvantaged students at school level.

The next two panels of Table III report the financial proficiency 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 on the baseline financial proficiency test. At baseline, students in schools implementing random and ability matching without differentiated instruction perform better on the course financial proficiency and financial knowledge measure than the control schools.

The last two panels of Table III report the results on the first and second post-treatment tests and provide a first assessment of the effectiveness of the financial education programme. For the first post-treatment test, students in treatment schools, in contrast to students in 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. We observe that the scores on the second post-treatment test are lower than on the first post-treatment test, suggesting that the positive effect of the programme declines over time. It should be noted that these first results should be interpreted with caution given the differences observed at baseline. Similar differences are found for both experimental studies separately in Table B.I in Appendix.

Table IV: Baseline Financial Scores

Dependent variable	Baseline Course Financial Proficiency		Baseline Financial Knowledge		Baseline Financial Behaviour	
Treat	0.135 (0.0839)	0.0469 (0.0817)	0.163** (0.0779)	0.0732 (0.0776)	0.0282 (0.0744)	-0.0132 (0.0737)
Treat * matching	-0.00341 (0.103)	0.0648 (0.0935)	0.00718 (0.0995)	0.0659 (0.0951)	-0.0201 (0.0874)	0.0267 (0.0823)
Treat * instruction	-0.213* (0.120)	0.0136 (0.0976)	-0.165 (0.115)	0.0130 (0.0989)	-0.172* (0.104)	0.00679 (0.0926)
<i>Controls</i>	No	Yes	No	Yes	No	Yes
<i>Obs.</i>	2,407	2,407	2,407	2,407	2,407	2,407
<i>p-value</i> Random match = Ability match & instructions	0.059	0.415	0.125	0.357	0.057	0.730

Note: Standard errors clustered at class level in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Controls: gender, SES, track, type of school, grade; *p-value* at the bottom is derived from the *F*-test comparing the estimate 'Treat' (*Random match*) with the combined estimate 'Treat + Treat * matching + Treat * instruction' (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation.

Table III indicates imbalances at baseline, which might be explained by the randomisation at school level. As a first intuitive test, we repeat the balance check at school level in Table B.II in Appendix. For all experimental conditions, we do not observe any significant differences at the five percent significance level compared to the control condition. More importantly, financial scores are found to be balanced.

As a second test to examine the extent to which the lack of balance in the experimental and control condition mirrors the subsequent estimated effects, we regress the baseline financial scores on a treatment dummy, interaction terms between the treatment dummy and dummies for the differentiation practices, and additional controls. Controlling for observed heterogeneity in school and student

background characteristics, we find the coefficients to be considerably small and insignificant for the overall baseline financial proficiency and the financial knowledge and behaviour measures separately, as presented in Table IV. The significant differences in pre-treatment test scores observed in Table III between different experimental conditions disappear when controlling for student background. These findings may serve as further evidence that the observed baseline imbalances do not affect the subsequent results.

3.2. Attrition

This section discusses the various levels at which attrition to the study occurred and the importance of it. Figure A.II and Figure A.III in Appendix show the attrition in respectively the first and second study. From 334 classes that 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). This source of attrition can partially be explained on practical grounds, i.e. as it was possible to implement the programme as a cross-curricular course, some teachers might have prioritised their own course material over the financial education programme, particularly in the second study, as it was organised 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 first 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. Examining the percentages of attrition within classes, we find that the absence of students during the completion of the post-test is the most plausible explanation for attrition, i.e. only 17.3 percent of the attrition is explained by teachers not adhering to the protocol. Given the high rates of attrition observed in both studies, we explore the consequences for the internal and external validity of our study in the following paragraphs.

We assess the internal validity via three analyses. First, we examine the baseline characteristics of all students who completed the pre-test (i.e. both compliers and non-compliers with the post-treatment test), as presented in Table B.III in Appendix. Although randomisation typically guarantees observed and unobserved characteristics to be equal at this stage, similar differences as in Table B.I are observed. One explanation for this finding is the high rate of attrition of schools prior to the completion of the pre-test, indicating the importance of assessing the external validity of our study.

Second, as the baseline characteristics were collected for all students, we test for selective attrition between the pre- and post-treatment test, as it can potentially bias our results. This is done by regressing an indicator for having missing data for the outcome measures on treatment indicators and controls. Table V shows that students from schools assigned to the experimental conditions were more likely to drop out in the first study. The finding remains significant when combining both studies. Teachers in the control condition received the didactic material conditional upon completion of the post-test and therefore, had a stronger incentive to make sure that this test was completed by the students.

This incentive was not present for teachers in the experimental conditions and can thus explain the higher attrition rates in the experimental conditions.

Table V: Selective Attrition

Dependent variable	Attrition Study I & II	Attrition Study I	Attrition Study II
Treat	0.159*** (0.0508)	0.174*** (0.0534)	0.122 (0.122)
Treat * matching	-0.0764 (0.0640)	-0.0663 (0.0702)	-0.117 (0.140)
Treat * instruction	0.136* (0.0717)	0.130 (0.0800)	0.148 (0.159)
<i>Controls</i>	Yes	Yes	Yes
<i>Obs.</i>	3,369	2,610	759
<i>p-value</i> Random match = <i>Ability match & instructions</i>	0.350	0.355	0.841

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); *p-value* at the bottom is derived from the *F*-test comparing the estimate 'Treat' (*Random match*) with the combined estimate 'Treat + Treat * matching + Treat * instruction' (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation.

Third, we compare non-complying (only completing the pre-test) and complying students (completing the post-test as well) within each condition and find significant differences for both studies, as presented in Table VI. In particular, for the control condition and condition of random matching, non-complying students are 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-test. These three attrition analyses demonstrate the importance of including control variables and assessing the robustness of the regression results (e.g., by matching techniques and bounding the estimates).

Finally, 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 of children with a mother having at most a secondary education degree, the percentage of children not speaking the official language (Dutch) at home, and the percentage of children receiving an allowance) to unique school identifiers participating in the study.¹¹ Table VII shows the sample of 107 schools that were willing to participate in the programme to be representative of the average Flemish school. 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. Yet, when complying and non-complying schools are compared, significant differences for two socioeconomic indicators (although only at the ten percent level) are

¹¹ AGODI, Cijfermateriaal - Leerlingenkenmerken (2017-2018), available at <http://www.agodi.be/cijfermateriaal-leerlingenkenmerken>.

observed. Students in non-complying schools are more likely to receive an allowance and have a mother with at most a secondary education degree.

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

STUDY I	Variables	Control	<i>p</i> - value	Random match	<i>p</i> - value	Ability match	<i>p</i> - value	Ability match & instruction	<i>p</i> - value
<i>Percentage of Attrition</i>		18.61%		28.19%		24.62%		40.18%	
<i>Number of Attrited Students</i>		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	<i>Academic</i>	-16.76%		-26.94%		-8.73%		-7.40%	
	<i>Technical</i>	12.54%	0.234	3.46%	0.009	3.88%	0.314	-2.48%	0.295
	<i>Vocational</i>	4.22%		23.40%		4.84%		9.88%	
<i>Background characteristics</i>									
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
Grade in mathematics (5)		0.15 (0.12)	0.216	-0.05 (0.13)	0.693	-0.13 (0.20)	0.508	-0.39 (0.16)	0.021
<i>Financial characteristics & scores</i>									
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	<i>p</i> - value	Random match	<i>p</i> - value	Ability match	<i>p</i> - value	Ability match & instruction	<i>p</i> - value
<i>Percentage of attrition</i>		23.45%		34.02%		22.92%		48.89%	
<i>Number of Attrited Students</i>		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	<i>Academic</i>	-67.52%		-42.06%		7.93%		21.45%	
	<i>Technical</i>	36.01%	0.041	37.85%	0.047	-8.29%	0.810	-37.35%	0.877
	<i>Vocational</i>	31.51%		4.08%		0.37%		15.9%	
<i>Background characteristics</i>									
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
Grade in mathematics (5)		0.39 (0.21)	0.093	-0.13 (0.22)	0.566	0.22 (0.28)	0.450	-0.31 (0.17)	0.093
<i>Financial characteristics & scores</i>									
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 track, relative differences are given.

Table VII: Comparison of Schools

Characteristic	Participating schools (complying & non-complying)	Total population of schools	<i>p</i>-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	<i>p</i>-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	<i>p</i>-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 that signed up for the programme, complying schools refer to schools where students completed both pre- and post-treatment tests.

4. Results

In this section, we discuss the estimates of the intent-to-treat analysis and examine heterogeneity in the treatment effect. In particular, we explore whether the observed findings vary along with a set of student and class characteristics, i.e. the financial proficiency, math ability, gender, socioeconomic status, language proficiency, and class size.

4.1. Intent-To-Treat analysis

The three first panels of Table VIII 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 treatment indicators and the pre-treatment value of the outcome, while the regression in the second column additionally controls for several school and student characteristics. Positive effects of the financial education programme for students' financial proficiency are found in both studies. As the coefficients for the course financial proficiency, financial knowledge, and behaviour of students obtained in the first study (panel 2 in Table VIII) and second study (panel 3 in Table VIII) are not significantly different from each other, as demonstrated by the *p*-values of Chi-square tests in panel 4, this suggests that the effect has external validity. Note that, given the similarities between both studies, we will only discuss the (more precise) pooled estimates of the composite sample in what follows.

The results suggest that, in the short term, the financial education programme is effective. In terms of magnitude, controlling for the observed heterogeneity, following the financial education programme

raises the performance of students in a financial proficiency test by 0.18 standard deviations on average. Intuitively, assuming a normal distribution, 0.18 standard deviations is the difference between a student at the 43rd percentile and the 50th percentile. The primary observation concerning the post-treatment performance of students is the improvement in financial knowledge (increase of 0.21 standard deviations on average). Financial behaviour of students improves by 0.07 standard deviations, though the estimate is not statistically significant. Further, in the second study, we measure financial attitudes of ninth grade students using three Likert-scale statements based on (i) the importance of financial literacy and (ii) saving, and (iii) the likelihood of the student comparing prices between shops before making a purchase. Table B.IV in Appendix suggests no improvement in students' financial attitudes after the financial education programme. The results for all outcome measures are robust against the inclusion of a set of control variables as reported in Table III.¹²

Ability matching and differentiated instruction appear, on average, unimportant for financial education. This holds for the course financial proficiency measure and, in particular, the financial knowledge of students. The financial knowledge increases, yet not statistically significant, by merely 0.06 standard deviations when students are matched to a peer with a similar ability level. Providing lower ability pairs additional instructions reduces students' average performance on the knowledge component by 0.04 standard deviations (again, the estimate is not statistically significant). Note that the absence of effects for the knowledge component confirms previous meta-analyses on ability matching in the wider educational science literature (Lou et al., 1996; Slavin, 1990). Per contra, we find that matching students according to ability does seem to increase the financial behaviour by 0.12 standard deviations on average in comparison to random matching, but this effect is significant only at the ten percent level. Moreover, adding differentiated instruction does not increase the effectiveness of ability matching for this outcome measure either. A similar pattern for financial attitudes can be observed, i.e. a positive effect of ability matching on the importance of financial literacy, whereas a negative influence of differentiated instruction on the importance of financial literacy and saving.

¹² In auxiliary regressions, we experimented with alternative control variables that capture additional student characteristics (i.e., the age, grade in mathematics and Dutch, and financial characteristics), peer and class characteristics (i.e., the class size, peers' pre-treatment score and grades in mathematics), and a teacher characteristic (i.e., years of teaching experience). The coefficients and interpretations of results remain robust against the inclusion of such controls.

Table VIII: Intent-to-Treat Analysis for Short-term Performance

Dependent variable		Course Financial Proficiency		Financial Knowledge		Financial Behaviour	
Study I & II (N = 2,407)	Treat	0.227*** (0.0704)	0.177** (0.0702)	0.254*** (0.0741)	0.206*** (0.0757)	0.115* (0.0627)	0.0671 (0.0600)
	Treat * matching	0.0780 (0.0885)	0.104 (0.0758)	0.0303 (0.0870)	0.0604 (0.0784)	0.105 (0.0793)	0.122* (0.0705)
	Treat * instruction	-0.179* (0.0973)	-0.0115 (0.0807)	-0.177* (0.0941)	-0.0407 (0.0805)	-0.129 (0.0902)	0.0357 (0.0788)
	<i>Controls</i> <i>p-value</i> Random match = Ability match & instructions	No 0.213	Yes 0.211	No 0.079	Yes 0.795	No 0.752	Yes 0.031
Study I (N = 1,896)	Treat	0.212*** (0.0805)	0.154* (0.0825)	0.222*** (0.0837)	0.152* (0.0880)	0.118 (0.0736)	0.0896 (0.0726)
	Treat * matching	0.147 (0.109)	0.149 (0.0925)	0.118 (0.105)	0.132 (0.0957)	0.134 (0.0939)	0.120 (0.0830)
	Treat * instruction	-0.207* (0.110)	-0.0550 (0.0880)	-0.206** (0.103)	-0.0901 (0.0866)	-0.123 (0.102)	0.0276 (0.0912)
	<i>Controls</i> <i>p-value</i> Random match = Ability match & instructions	No 0.477	Yes 0.208	No 0.301	Yes 0.599	No 0.900	Yes 0.063
Study II (N = 511)	Treat	0.256* (0.145)	0.209 (0.143)	0.346** (0.159)	0.297* (0.150)	0.0781 (0.101)	0.0336 (0.103)
	Treat * matching	-0.0758 (0.149)	0.000759 (0.138)	-0.181 (0.148)	-0.0957 (0.134)	0.0307 (0.151)	0.0865 (0.140)
	Treat * instruction	-0.157 (0.252)	0.199 (0.181)	-0.175 (0.263)	0.128 (0.185)	-0.210 (0.205)	0.185 (0.170)
	<i>Controls</i> <i>p-value</i> Random match = Ability match & instructions	No 0.362	Yes 0.338	No 0.187	Yes 0.878	No 0.340	Yes 0.125
<i>p-value</i> Chi-square test	Treat		0.732		0.399		0.651
	Treat * matching		0.363		0.160		0.832
	Treat * instruction		0.198		0.276		0.405

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); *p*-value at the bottom of each panel is derived from the *F*-test comparing the estimate ‘Treat’ (*Random match*) with the combined estimate ‘Treat + Treat * matching + Treat * instruction’ (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation; *p*-values of the Chi-square test statistic are calculated to examine whether the treatment coefficients significantly differ between Study I and Study II.

Further, to investigate whether the financial knowledge of students is retained after time has elapsed, we look at students’ performance in the second post-treatment test, administered approximately six weeks after the lectures as a homework. Although we must carefully interpret the results due to the smaller and rather selective samples for the experimental conditions, students for whom ability matching and differentiated instruction was implemented are found to perform better in the second post-treatment financial proficiency test in comparison to the other conditions, as presented in Table IX.¹³ Interestingly, the positive effect of the financial education programme persists only for the knowledge component.

¹³ Only 23 percent of students in the experimental conditions completed the second post-treatment test. To examine whether the sample is representative, we repeat the attrition analysis for the second-post treatment test in Appendix D.

Table IX: Intent-to-Treat Analysis for Long-term Performance

Dependent variable	Course Financial Proficiency		Financial Knowledge		Financial Behaviour	
Treat	0.180 (0.154)	0.196 (0.179)	0.0561 (0.127)	0.0965 (0.145)	0.253 (0.155)	0.231 (0.169)
Treat * matching	-0.197 (0.178)	-0.244 (0.201)	-0.153 (0.141)	-0.199 (0.156)	-0.0869 (0.189)	-0.132 (0.200)
Treat * instruction	0.0695 (0.177)	0.321* (0.172)	0.135 (0.161)	0.366** (0.167)	-0.109 (0.158)	0.0829 (0.139)
<i>Controls</i>	No	Yes	No	Yes	No	Yes
<i>Obs.</i>	1,208	1,208	1,208	1,208	1,208	1,208
<i>p-value</i> Random match = Ability match & instructions	0.528	0.715	0.923	0.359	0.273	0.794

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 tests; *p*-value at the bottom is derived from the *F*-test comparing the estimate ‘Treat’ (*Random match*) with the combined estimate ‘Treat + Treat * matching + Treat * instruction’ (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation; For students in the control condition, the score on the first post-test is used to measure long-term performance.

4.2. Heterogeneous effects

Let us first consider the heterogeneity in financial proficiency scores. To set the scope, we illustrate in Figure A.IV in Appendix that the performance of students improves across the full distribution of post-treatment financial proficiency scores. In particular, we observe a rightward shift in the scores of students in all three experimental conditions compared to students in the control condition, differences that are statistically significant as measured by the Kolmogorov-Smirnov test.¹⁴

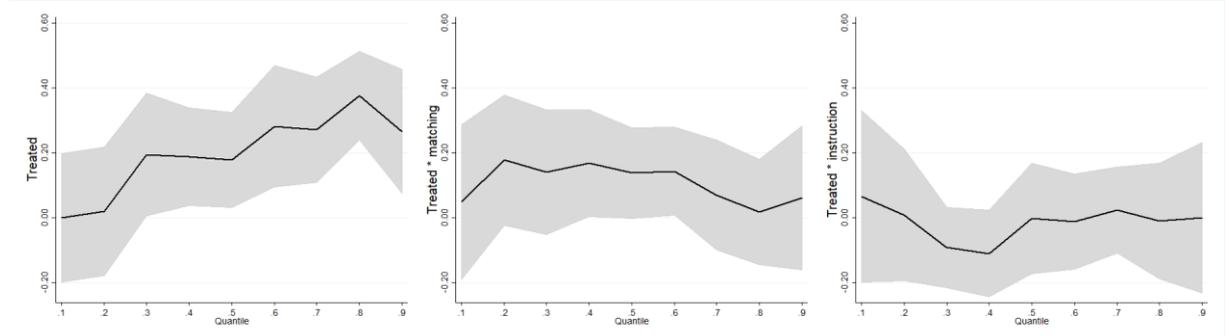


Figure I: Quantile Regression for Course Financial Proficiency

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

We further analyse the heterogeneity of the treatment effects by estimating a quantile regression for each decile of the financial proficiency measure. Apart from the first decile, Figure I indicates significant treatment effects throughout. However, we find the distribution-wide effects to be conditional upon the

¹⁴ 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 is compared with the distribution of the control condition. The tests reject equality of distributions for all three experimental conditions ($p = 0.000$, $p = 0.000$ and $p = 0.076$).

teaching delivery method. Despite the wide confidence intervals, Figure I suggests that low-achieving students in the financial proficiency test benefit whenever they are matched to a peer with a similar ability level, whereas high-achieving students perform equally well in randomly-matched pairs. For all deciles, differentiating instructions in the classroom does not seem to significantly improve the performance of ability-matched student pairs.

Second, as differentiated instruction was given based on students' grade in mathematics, and not students' financial proficiency, the former quantile regression does not provide information on the effects of differentiation for those students who were actually eligible for additional instructions. Using students' self-assessed grade in mathematics from the baseline financial proficiency test, we split students in each class into three groups in order to identify low, medium, and high ability students.¹⁵ Low ability students received additional instructions in all modules, medium ability students in the first module of the material only, and high ability students did not receive additional instructions. The fourth column of Table X suggests no heterogeneous treatment effect of the programme for students with different abilities. Moreover, although expected from the educational science literature (e.g. Lou et al., 1996), pairing students according to ability does not seem to significantly benefit (or harm) low, medium, and high ability students, nor does differentiating the instructions.

Table X: Heterogeneous Effects by Math Ability

Dependent variable: Course Financial Proficiency	Math ability			Difference <i>p</i> -value
	<i>Low</i>	<i>Medium</i>	<i>High</i>	
Treat	0.151 (0.102)	0.219** (0.110)	0.158* (0.0948)	0.837
Treat * matching	0.106 (0.114)	0.0425 (0.125)	0.147* (0.0831)	0.796
Treat * instruction	-0.00537 (0.118)	0.0330 (0.127)	-0.0601 (0.103)	0.159
<i>Controls</i>	Yes	Yes	Yes	
<i>Obs.</i>	739	799	869	
<i>p</i> -value Random match = Ability match & instructions	0.348	0.543	0.423	

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, track, type of school, grade, SES; *p*-value at the bottom is derived from the *F*-test comparing the estimate 'Treat' (*Random match*) with the combined estimate 'Treat + Treat * matching + Treat * instruction' (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation.

Third, given a large body of research on financial literacy has identified differences in financial literacy levels by SES and gender (e.g., Lusardi, Mitchell, & Curto, 2010), we explore whether our financial education programme has heterogeneous effects with respect to these characteristics. Table

¹⁵ Given that students assessed their performance in mathematics on a five-point scale, group sizes might be unequally distributed, which we then account for by randomly splitting students to the closest ability groups (note that redrawing the students delivers robust outcomes).

B.V in Appendix provides the estimates from the subgroup analyses and does not reveal significant treatment heterogeneity.

Table XI: Heterogeneous Effects by Class size

Dependent variable: Course Financial Proficiency	Class size		Difference <i>p</i> -value
	<i>Small/average</i>	<i>Large</i>	
Treat	0.244** (0.0946)	0.136 (0.104)	0.437
Treat * matching	-0.0604 (0.0964)	0.218* (0.111)	0.056
Treat * instruction	0.160 (0.101)	-0.149 (0.118)	0.044
<i>Controls</i>	Yes	Yes	
<i>Obs.</i>	1,195	1,212	
<i>p</i> -value Random match = Ability match & instructions	0.323	0.515	

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, track, type of school, grade, SES; Large classes count 19 students or more; *p*-value at the bottom is derived from the *F*-test comparing the estimate 'Treat' (*Random match*) with the combined estimate 'Treat + Treat * matching + Treat * instruction' (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation.

Fourth, following Barrow, Markman, and Rouse (2009) and Lou et al. (1996), we expect students in larger classrooms to significantly benefit from differentiation for two reasons. First, teachers in large classes are required to divide the individual instruction time among more students, and larger classes are subjected to higher disruption levels (Lazear, 2001). Second, it is likely that students' performances are more heterogeneous within larger classes. Accordingly, since it is more difficult in a traditional classroom setting to design uniform teaching material appropriate for *all* students, differentiated instruction is expected to have the largest impact on students in larger classes. Though we expect a larger isolated effect of ability matching as well, the sign of the effect is *ex ante* unknown, i.e. despite student variability in the classroom is reduced by matching, we expect that low-ability students may be particularly harmed by it as they can no longer benefit from their higher ability peers.

Table XI shows the results by class size. The subgroup analysis for large classes (defined as a class with 19 students or more) reveals a positive impact of ability matching as compared to random matching, i.e. the performance of students in large classes improves by 0.22 standard deviations, respectively. The hypothesis of increased performance by differentiating instructions, however, is not confirmed in our data. For students in smaller classes, ability matching and differentiated instruction seem trivial within the context of financial literacy education.

Finally, the language proficiency of students can be considered as an alternative mediating factor while assessing the effectiveness of differentiated instruction since non-native students in our study are less proficient in the teaching language.¹⁶ Given our programme design led students independently through the different steps via the information booklet, in particular for the first module of the material,

¹⁶ On average, non-native students assess their grade in Dutch on a five point scale 0.73 points lower than native students.

we assume non-native students had more difficulties in reading and understanding all information in the booklet (Figlio, Rush, & Lu, 2013). Accordingly, using additional hints indicating where to find the correct answer to a question in the information booklet, we expect differentiated instruction to mitigate the language disadvantage, enabling non-native students to work at their own ‘language-level’.

This hypothesis is confirmed by our data, as illustrated in Table XII. The p -values from the tests comparing random matching to ability matching combined with differentiated instruction indicate that students who do not speak the teaching language at home significantly benefit from differentiated instructions during the lectures. Particularly, ability matching seems to outperform random matching for these students, only if additional instructions are provided ($p = 0.014$). This finding is consistent with the educational science literature that emphasizes the crucial role of adapting instructions when matching students according to ability (Lou et al., 1996; Slavin, 1990). The performance of native students, on the other hand, remains similar regardless of the type of matching or level of instructions.

Table XII: Heterogeneous Effects by Language Proficiency

Dependent variable: Course Financial Proficiency	Language Proficiency		Difference p -value
	<i>Non-native</i>	<i>Native</i>	
Treat	0.220* (0.117)	0.178** (0.0758)	0.735
Treat * matching	0.148 (0.160)	0.0814 (0.0820)	0.698
Treat * instruction	0.110 (0.161)	-0.0812 (0.0869)	0.251
<i>Controls</i>	Yes	Yes	
<i>Obs.</i>	485	1,922	
p -value Random match = Ability match & instructions	0.014	0.998	

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, track, type of school, grade, number of holidays per year; p -value at the bottom is derived from the F -test comparing the estimate ‘Treat’ (*Random match*) with the combined estimate ‘Treat + Treat * matching + Treat * instruction’ (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation.

5. Robustness analysis

We assess the robustness of our main results with respect to non-compliance, attrition and imbalances in covariates.

5.1. Non-compliance and Instrumental variables

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, we control for it by using the assigned treatment allocations as instruments for the actual implementation

of student matching and differentiated instruction (Athey & Imbens, 2017). The programme compliance is measured through teacher surveys, administered after the lectures, in which teachers were asked whether they correctly grouped the students (either randomly or according to ability) and gave the 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:

$$1^{\text{st}} \text{ stage: } \textit{Treat}_{cij} = \mu + \gamma'_0 \textit{Treat}_{is} + \gamma_1 y_{ijs}^0 + \sum \gamma'_2 S_s + \sum \gamma'_3 X_i + \delta_{ij} \quad (2)$$

$$\textit{matching}_{cij} = \rho + \lambda'_0 \textit{matching}_{is} + \lambda_1 y_{ijs}^0 + \sum \lambda'_2 S_s + \sum \lambda'_3 X_i + v_{ij} \quad (3)$$

$$\textit{instruction}_{cij} = \kappa + \theta'_0 \textit{instruction}_{is} + \theta_1 y_{ijs}^0 + \sum \theta'_2 S_s + \sum \theta'_3 X_i + v_{ij} \quad (4)$$

$$2^{\text{nd}} \text{ stage: } y_{ijs}^1 = \alpha + \beta_0 \widehat{\textit{Treat}}_{cij} + \beta_1 \widehat{\textit{Treat}}_{cij} * \widehat{\textit{matching}}_{cij} + \beta_2 \widehat{\textit{Treat}}_{cij} * \widehat{\textit{instruction}}_{cij} + \beta_3 y_{ijs}^0 + \sum \beta'_4 S_s + \sum \beta'_5 X_i + \epsilon_{ij} \quad (5)$$

where $\widehat{\textit{Treat}}_{cij}$, $\widehat{\textit{matching}}_{cij}$, and $\widehat{\textit{instruction}}_{cij}$ denote dummies indicating whether the teacher of student i in class j followed the instructions correctly, i.e. formed student pairs and/or gave differentiated instruction as advised.

Table B.VII in Appendix shows that the treatment effects are similar as in the main specification, but are less precisely estimated. The interpretation of the main results remains valid, i.e. no differences are found between random and ability matching. Differentiating instructions does not seem to improve students' performance. Moreover, it should be noted that the positive effect of ability matching for the financial behaviour measure disappears when accounting for teacher imperfect compliance.¹⁷

5.2. 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 affects the results of this paper, estimates are bounded using the bounding approach proposed by Lee (2009). Two identifying assumptions are required for Lee bounds, i.e. a random assignment to conditions and a monotonicity condition such that the treatment assignment affects attrition in one direction only. Both assumptions, a priori, hold in the two 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 A.V 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. Figures A.V, A.VI, and A.VII in Appendix present the Lee bounds for all outcome measures for the composite sample, study I, and study II, respectively. The figures show that the main intent-to-treat estimates are well within the bounds. In Figure A.V, we

¹⁷ The same interpretation remains valid when excluding the non-complying classes from the intent-to-treat regression analysis.

observe that at a trimming portion of 35 percent of what should be trimmed to equalize response rates across conditions, the confidence intervals of the intent-to-treat estimate and the upper and lower bounds for the average treatment effect overlap and hence, our estimate is robust at this stage. Furthermore, the estimates for treatment heterogeneity by ability matching and differentiated instruction appear robust at a trimming portion of 100 percent and 75 percent, respectively.

5.3. Baseline imbalance and 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 Blackwell, Iacus, King, and Porro (2009). CEM enables the generation of covariate balance *ex ante* and the comparison of learning outcomes of similar students in the control and experimental conditions. In particular, covariates are 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 conditions. After unmatched observations are deleted, the original uncoarsened values of covariates are retained and main intent-to-treat models are 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 B.VIII, 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 and this 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 differences in the proficiency or ability levels of students. This paper conducted two randomised control trials measuring the impact of two differentiation practices, i.e. ability matching and differentiated instruction, on the effectiveness of a financial education programme for eighth and ninth grade students in Flemish secondary schools.

The results showed that students’ financial proficiency improved by 0.18 standard deviations immediately after the programme. The improvement in financial proficiency was mainly driven by an increase in financial knowledge. Students’ financial behaviour appeared unaffected by the programme. Moreover, though we were unable to provide evidence on actual financial decision-making, self-reported measures of financial attitudes for ninth grade students revealed no significant changes either. On average, considering all types of students, we did not observe general effects of ability matching and differentiated instruction for students’ short-term financial proficiency and financial knowledge, and

financial behaviour when accounting for teacher imperfect compliance. However, approximately six weeks after the lectures, students for whom ability matching and differentiated instruction was implemented were found to retain the increase in financial knowledge as compared to the other experimental conditions. Note that this result should be interpreted with caution given the selectivity and size of the sample, and the homework format of the test. Further, we acknowledge that this finding as well represents a rather short-term effect and that ultimately, longer-term follow-ups would be needed to understand the full effect of the programme, especially for financial behaviour.

To add some perspective to these findings, it is worth noting that in the large-scale school-based financial education evaluation by Bruhn et al. (2016) the overall financial proficiency of students increased by a quarter of a standard deviation, while other experimental studies indicate more modest improvements, around 0.15 standard deviations (Frisancho, 2018; Villanueva et al., 2018). As such, our observed effect of 0.18 standard deviations is in line with earlier literature. Further, despite measuring changes in actual financial decision-making was beyond the scope of this paper and has been found challenging in the literature (Fernandes, Lynch, & Netemeyer, 2014), we would like to point out the study by Bruhn and colleagues (2016) where students' financial decision-making significantly changed after the programme.

We explored several factors by which the differentiation practices could have affected the programme effectiveness. Pairing students according to ability did not seem to significantly benefit (or harm) low, medium, and high ability students, nor did differentiating the instructions. However, we observed that providing additional instructions to lower ability students did benefit those students who do not speak the official teaching language at home. Dutch speaking students performed equally well regardless of the type of matching or level of instructions. As non-native students in our sample were less proficient in the teaching language, we argue that the differentiated instruction in the programme permitted students to follow the lectures at their own language-level (Figlio, Rush, & Lu, 2013). Further, ability matching had a positive impact on students in larger classes, whereas adding differentiated instruction did not increase their performances. Despite this finding is in contrast to previous research (Barrow, Markman, & Rouse, 2009), it is consistent with the following explanations. First, while the programme was designed to deliberately minimise the teacher role, teachers might still have interacted with their students. If they did not follow instructions as advised and helped students with the content of the material, the effect of differentiated instruction was likely to be mitigated. Via a descriptive analysis of the classroom observations, the occurrence of such events was reported for the majority of classroom observations in all conditions. Second, it is worth noting that due to the selective secondary education system in Flanders, students' abilities within classrooms were already relatively similar. Therefore, we may suggest that differentiated instruction implemented in a 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, 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 by showing that a relatively small intervention substantially increases the financial proficiency of secondary school students. Second, 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 raised the learning outcomes of non-native students in classrooms without deteriorating the performance of their peers. Given that non-native students are often found to sort themselves into schools with more academically disadvantaged native students (Geay et al., 2013), computer-based differentiation may be particularly beneficial for such schools. Accordingly, the practice of ability matching combined with differentiated instruction deserves the attention of the policymakers as it may be more cost-effective and easier to implement than other educational interventions.

To conclude, our study provides several avenues for further research. First, given that we were unable to identify the exact pairing of students, and hence, average effects of random and ability matching in the classroom were estimated, it may be particularly interesting to explore underlying peer effects. Second, the programme was designed in such a way that students in one class received additional instructions whereas if they had been in a class with lower ability students they wouldn't have received them (as they would then belong to the better-performing students). Accordingly, the evaluation of a differentiated approach where the material is adapted to the specific, individual needs of students should be considered for future work.

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*.
- Athey, S., & Imbens, G. (2017). The Econometrics of Randomised Experiments. *Handbook of Economic Field Experiments*, 1:73-140.
- Barrow, L., Markman, L., & Rouse, C. E. (2009). Technology's Edge: The Educational Benefits of Computer-Aided Instruction. *American Economic Journal: Economic Policy*, 1(1), 52–74.
- 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.
- 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.
- De Witte, K., Holz, O., & De Beckker, K. (2019). *Financial Education - Current practices and future challenges*. Waxmann Verlag. pp. 180. ISBN 978-3-8309-9063-5.
- 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.
- Epple, D., Newlon, E., & Romano, R. (2002). Ability tracking, school competition, and the distribution of educational benefits. *Journal of Public Economics*, 83, 1–48.
- Fernandes, D., Lynch, J. G., & Netemeyer, R. G. (2014). Financial Literacy, Financial Education, and Downstream Financial Behaviors. *Management Science*, 60(8), 1861–1883.
- 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*.
- Geay, C., McNally, S., & Telhaj, S. (2013). Non-native speakers of English in the classroom: What are the effects on pupil performance? *The Economic Journal*, 123(August), 281–307.
- 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.
- Lazear, E. P. (2001). Education Production. *The Quarterly Journal of Economics*, CXVI(3), 777–803.
- 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., & Curto, V. (2010). Financial Literacy among the Young. *The Journal of Consumer Affairs*, 44(2), 358–380.
- Lusardi, A., & Mitchell, O. S. (2014). The Economic importance of Financial Literacy: Theory and Evidence. *Journal Of Economic Literature*, 52(1), 5–44.
- Maldonado, J. E., De Witte, K., & Declercq, K. (2019). *The Effects of Parental Involvement in Homework – Two Randomised Controlled Trials in Financial Education*. Unpublished results, pp. 52.
- Mandell, L. (2008). *The Financial Literacy of Young American Adults: Results of the 2008 National Jump\$trats 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. (2016). *PISA 2015 Financial Literacy Framework*. PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy, OECD Publishing, Paris.
- OECD. (2017). *PISA 2015 Results (Volume IV): Students' Financial Literacy*. PISA, OECD Publishing, Paris.
- Slavin, R. E. (1990). Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis. *Review Of Educational Research*, 60(3), 471–499.
- van Rooij, M. C. J., Lusardi, A., & Alessie, R. J. M. (2012). Financial Literacy, Retirement Planning and Household Wealth. *The Economic Journal*, 122, 449–478.
- 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 A: Figures

Hint!
Look it up on information sheet 7.



Key 6	True	False
If you pay contactless with your debit card or smartphone, you can pay up to € 50,00 without using your PIN.	_ .	_ _ .

Key 7



Calculation 1

You go abroad to study for 1 year. Your parents give you a prepaid card. The card costs € 15,00 per year. Every month, your parents transfer an amount to your account (for free). You withdraw this amount every month. This costs you € 4,00 per withdrawal. Every two months, you consult your balance and you pay € 0,50 per text message. How much will your prepaid card cost your parents after 1 year?

Always use two digits after the comma!

7	8	9
4	5	6
1	2	3
0	,	

Tip!

1. Calculate the cost of withdrawing the money (= cost of monthly withdrawal x 12)
2. Calculate the cost of consulting the balance (= cost of consulting every 2 months x 6)
3. The total cost = cost 1 + cost 2 + yearly cost

KU LEUVEN

wikifin.be



Figure A.I: Differentiated Instruction in Didactic Material

Note: Sticky notes in didactic material indicate where to look up the information in the booklet and how to do the calculation needed to solve the problem.

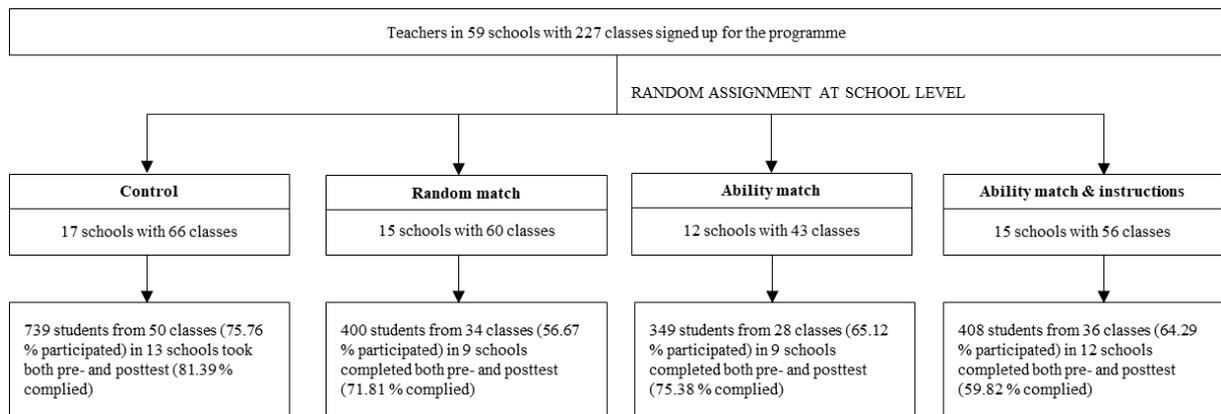


Figure A.II: Selection of Participants per Condition for Study I

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

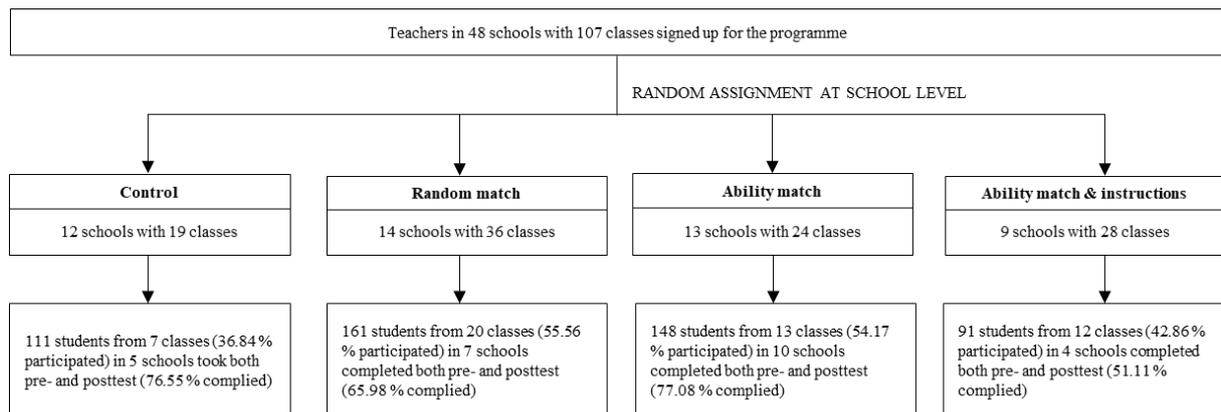


Figure A.III: Selection of Participants per Condition for Study II

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

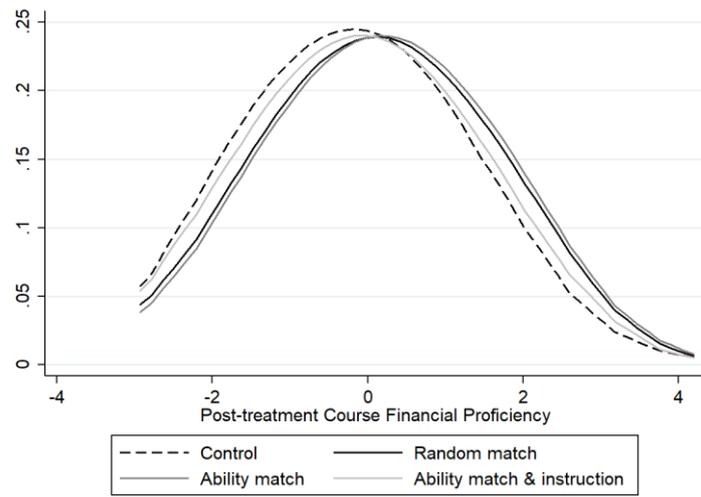


Figure A.IV: Distribution Shift of Course Financial Proficiency

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

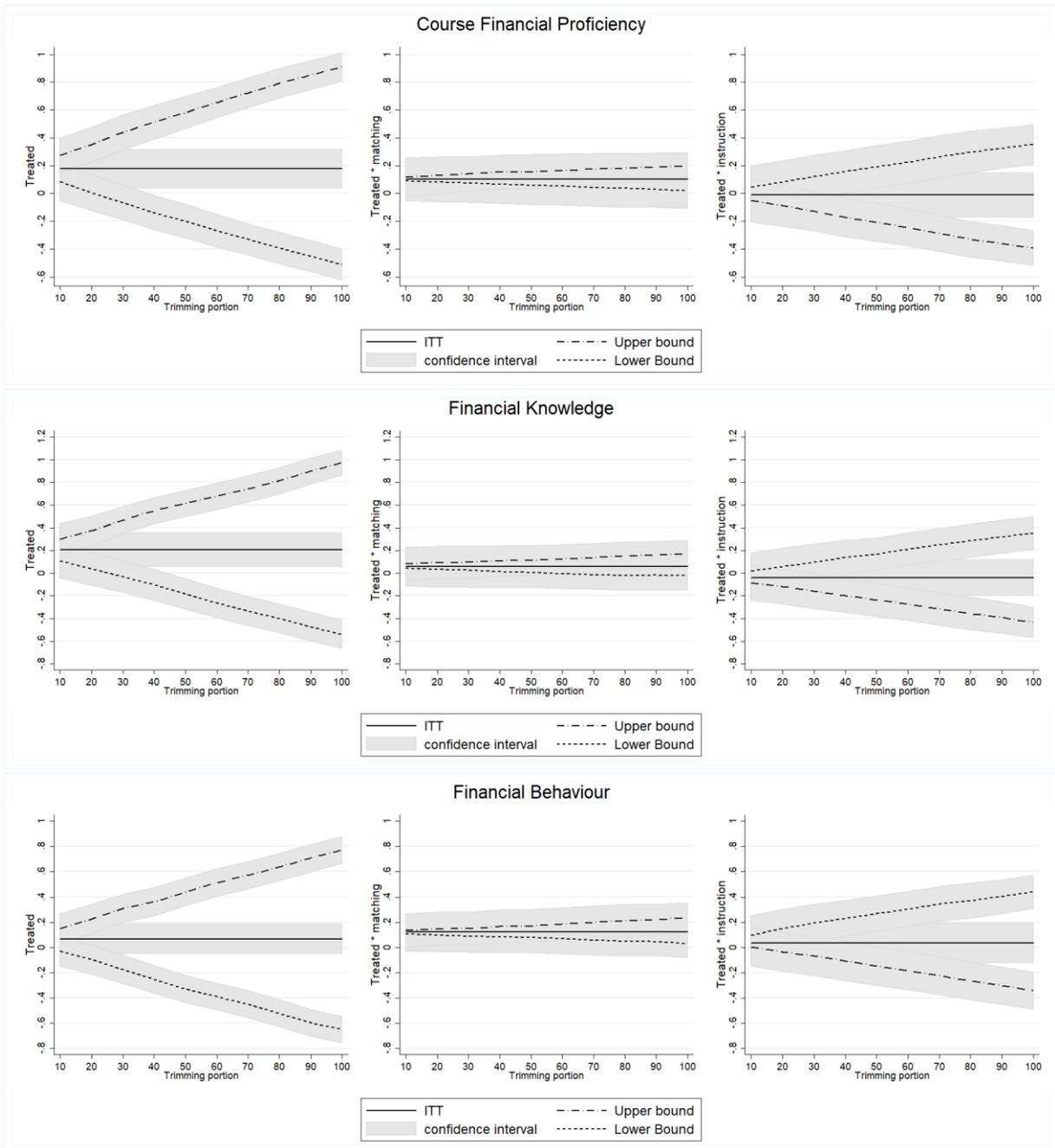


Figure A.V: Lee Bounds for Study I & II

Note: The approach described in Merlino, Steinhardt, and Wren-Lewis (2018) is followed when computing 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 (Random match and Ability match), the samples are trimmed by dropping students with the smallest (highest) residuals. The bounds estimates are then derived from a 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.

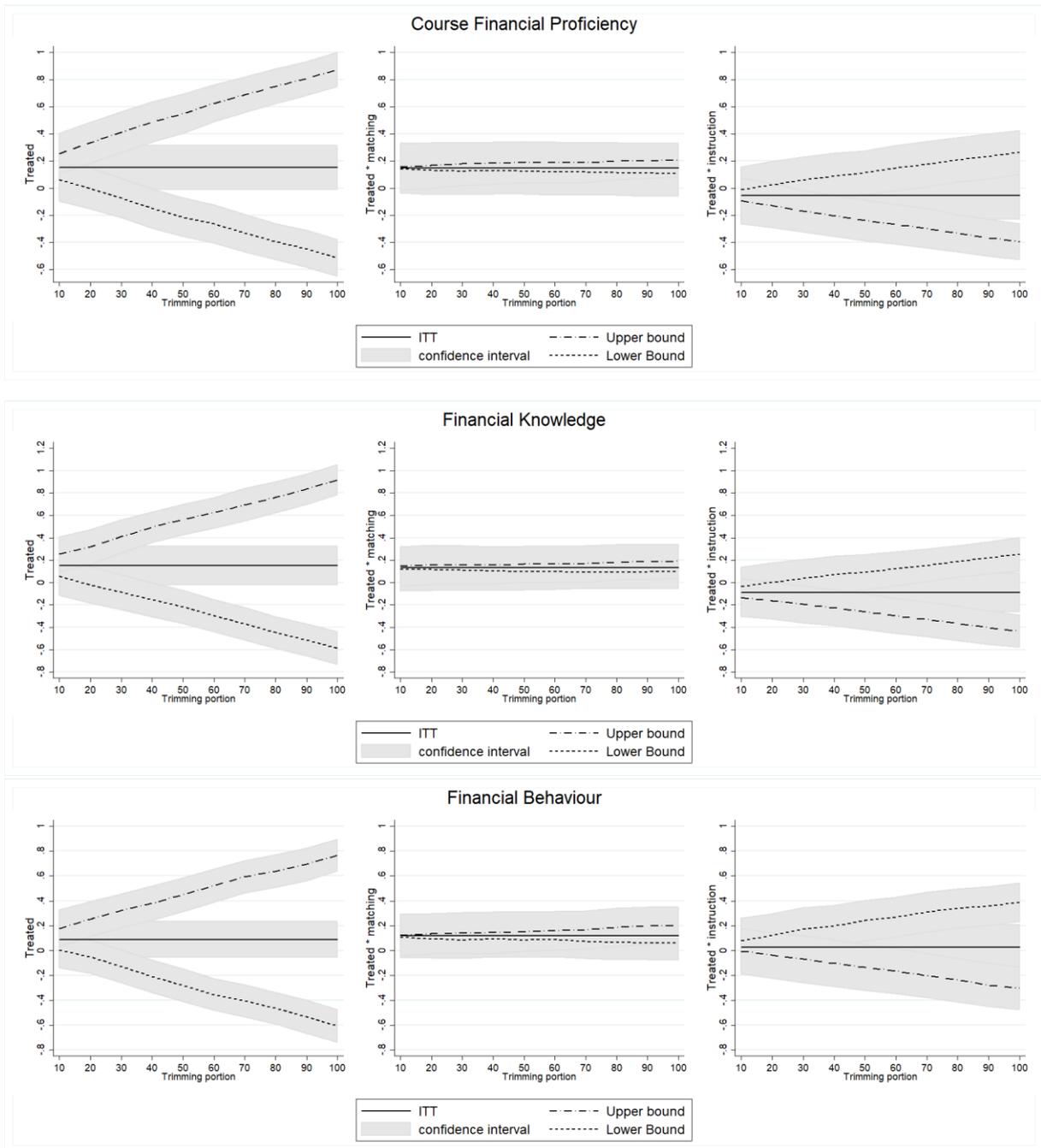


Figure A.VI: Lee Bounds for Study I

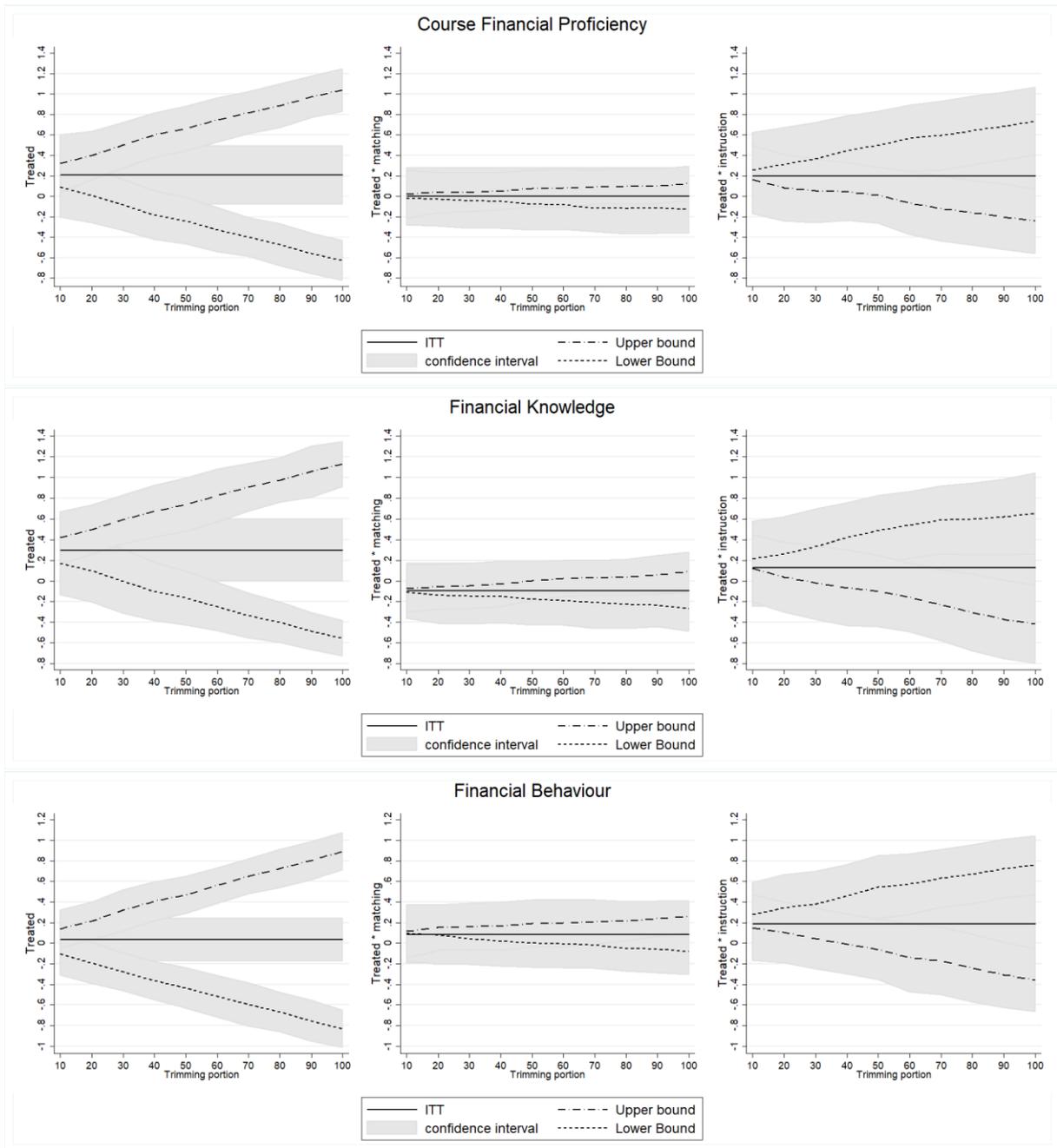


Figure A.VII: Lee Bounds for Study II

Appendix B: Tables

Table B.I: Descriptive Statistics per Study

STUDY I	Variables	Control	Random match	<i>p</i> - value	Ability match	<i>p</i> - value	Ability match & instruction	<i>p</i> - value
	<i>Number of Schools</i>	13	9		9		12	
	<i>Number of Students</i>	739	400		349		408	
	Type of school (private)	0.68	0.98	0.000	0.92	0.008	0.89	0.017
	Class size	18.61 (4.70)	17.37 (4.78)	0.221	17.38 (4.68)	0.210	16.73 (5.15)	0.070
	Track							
	<i>Academic</i>	653 (88.4%)	360 (90.0%)		306 (87.68%)		261 (63.97%)	
	<i>Technical</i>	21 (2.84%)	32 (6.20%)	0.341	14 (4.01%)	0.987	95 (23.28%)	0.048
	<i>Vocational</i>	65 (8.80%)	8 (1.55%)		29 (8.31%)		52 (12.75%)	
	<i>Background characteristics</i>							
	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
	Grade in mathematics (5)	3.10 (1.31)	3.08 (1.25)	0.873	3.31 (1.15)	0.125	2.92 (1.33)	0.277
	<i>Financial characteristics</i>							
	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
	<i>Pre-treatment financial scores</i>							
	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
	<i>1st post-treatment financial scores</i>							
	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
	<i>2nd post-treatment financial scores</i>							
	Course financial proficiency (9)		3.04 (1.40)		3.25 (1.57)		2.94 (1.63)	
	Financial knowledge (6)		1.93 (1.12)		1.94 (1.16)		1.94 (1.21)	
	Financial behaviour (3)		1.11 (0.74)		1.31 (0.78)		1.01 (0.78)	
STUDY II	Variables	Control	Random match	<i>p</i> - value	Ability match	<i>p</i> - value	Ability match & instruction	<i>p</i> - value
	<i>Number of Schools</i>	5	7		10		4	
	<i>Number of Students</i>	111	161		148		91	
	Type of school (private)	0.92	0.95	0.731	0.92	1.00	0.82	0.534
	Class size	17.64 (4.08)	12.10 (6.08)	0.022	16.35 (6.04)	0.613	10.96 (2.99)	0.001
	Track							
	<i>Academic</i>	88 (79.28%)	149 (92.55%)		106 (71.62%)		15 (16.48%)	
	<i>Technical</i>	14 (12.61%)	7 (4.35%)	0.400	23 (15.54%)	0.752	19 (20.88%)	0.003
	<i>Vocational</i>	9 (8.11%)	5 (3.11%)		19 (12.84%)		57 (62.64%)	
	<i>Background characteristics</i>							
	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
	Grade in mathematics (5)	2.82 (1.11)	2.98 (1.29)	0.538	3.05 (1.07)	0.253	3.26 (1.18)	0.108
	<i>Financial characteristics</i>							
	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	3.47 (1.47)	0.812	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

1st 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

2nd post-treatment financial scores

Course financial proficiency (9)		3.48 (1.67)		2.86 (1.82)		3.96 (1.55)	
Financial knowledge (6)		2.04 (1.19)		1.86 (1.27)		2.61 (1.31)	
Financial behaviour (3)		1.44 (0.70)		1.00 (0.79)		1.35 (0.78)	

Note: Mean values and standard deviation in parentheses; For track, absolute and relative frequencies are given; *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 standardisation; Students from control schools did not complete the second post-treatment test. Accordingly, their first post-treatment test scores will be used as a comparison in the analysis of the second post-test, considering no learning effects of the test itself.

Table B.II: Descriptive Statistics Study I & II at School Level

Variables	Control	Random match	<i>p</i> -value	Ability match	<i>p</i> -value	Ability match & instruction	<i>p</i> -value
<i>Number of Schools</i>	17	16		17		15	
Type of school (private)	0.67	0.81	0.351	0.72	0.727	0.67	0.879
Class size	16.63 (5.77)	12.73 (5.51)	0.054	14.18 (5.12)	0.187	13.84 (5.95)	0.211
Track	1.44 (0.69)	1.45 (0.68)	0.950	1.57 (0.77)	0.586	1.62 (0.64)	0.551
Background characteristics							
Gender (female)	0.53	0.56	0.599	0.45	0.291	0.43	0.183
Age (years)	14.44 (0.54)	14.60 (0.54)	0.393	14.86 (0.75)	0.059	14.59 (0.62)	0.473
Number of holidays per year (4)	2.91 (0.30)	2.85 (0.41)	0.662	2.90 (0.41)	0.970	2.93 (0.45)	0.725
Language (Dutch)	0.80	0.84	0.657	0.85	0.476	0.63	0.073
Grade in mathematics (5)	3.05 (0.49)	3.03 (0.50)	0.908	3.02 (0.60)	0.836	3.07 (0.78)	0.843
Financial characteristics							
Importance of financial literacy (5)	4.23 (0.16)	4.17 (0.30)	0.499	4.21 (0.24)	0.878	4.14 (0.25)	0.300
Importance of saving (5)	4.16 (0.26)	4.15 (0.36)	0.933	4.20 (0.27)	0.602	4.12 (0.38)	0.802
Comparison shops before purchase (5)	3.48 (0.48)	3.39 (0.59)	0.635	3.51 (0.41)	0.835	3.61 (0.23)	0.259
Pre-treatment financial scores							
Course financial proficiency (9)	2.95 (0.72)	3.11 (0.71)	0.522	3.17 (0.88)	0.409	2.98 (0.78)	0.749
Financial knowledge (6)	1.93 (0.43)	1.89 (0.50)	0.827	2.03 (0.66)	0.357	1.84 (0.45)	0.824
Financial behaviour (3)	1.02 (0.30)	1.22 (0.30)	0.291	1.14 (0.33)	0.660	1.14 (0.40)	0.711

Note: Mean values and standard deviation in parentheses; *p*-values are derived from a difference in means test between students in the treatment and control schools via a regression of the characteristic on the treatment indicator.

Table B.III: Baseline Characteristics of Students per Study

STUDY I	Variables	Control	Random match	<i>p</i> - value	Ability match	<i>p</i> - value	Ability match & instruction	<i>p</i> - value
	<i>Number of Schools</i>	16	11		10		14	
	<i>Number of Students</i>	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							
	<i>Academic</i>	774 (85.24%)	459 (82.41%)		396 (85.53%)		416 (61.00%)	
	<i>Technical</i>	47 (5.18%)	50 (8.98%)	0.862	23 (4.97%)	0.996	152 (22.29%)	0.014
	<i>Vocational</i>	87 (9.58%)	48 (8.62%)		44 (9.50%)		114 (16.72%)	
	<i>Background characteristics</i>							
	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
	Grade in mathematics (5)	3.13 (1.30)	3.06 (1.27)	0.626	3.28 (1.17)	0.258	2.77 (1.35)	0.010
	<i>Financial characteristics & scores</i>							
	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	Random match	<i>p</i> - value	Ability match	<i>p</i> - value	Ability match & instruction	<i>p</i> - value
	<i>Number of Schools</i>	7	10		9		6	
	<i>Number of Students</i>	145	244		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							
	<i>Academic</i>	92 (63.45%)	191 (78.28%)		141 (73.44%)		48 (26.97%)	
	<i>Technical</i>	24 (16.55%)	39 (15.98%)	0.320	21 (10.94%)	0.670	79 (44.38%)	0.179
	<i>Vocational</i>	29 (20.00%)	14 (5.74%)		30 (15.63%)		51 (28.65%)	
	<i>Background characteristics</i>							
	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
	Grade in mathematics (5)	2.91 (1.14)	2.94 (1.29)	0.895	3.10 (1.08)	0.327	3.11 (1.23)	0.315
	<i>Financial characteristics & scores</i>							
	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 deviation in parentheses; For track, absolute and relative frequencies are given; *p*-values are derived from a 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 standardisation.

Table B.IV: Financial Attitudes of Ninth Grade Students

Dependent variable	Financial attitudes					
	Importance financial literacy		Importance saving		Comparison shops before purchase	
Treat	-0.309 (0.320)	-0.387 (0.280)	-0.00672 (0.207)	-0.0119 (0.210)	-0.143 (0.186)	-0.145 (0.196)
Treat * matching	0.675** (0.304)	0.796** (0.328)	0.235 (0.235)	0.267 (0.243)	-0.180 (0.216)	-0.146 (0.209)
Treat * instruction	-1.108*** (0.274)	-0.987*** (0.291)	-0.632* (0.328)	-0.619* (0.354)	-0.243 (0.231)	-0.0763 (0.277)
<i>Controls</i>	No	Yes	No	Yes	No	Yes
<i>Obs.</i>	504	504	504	504	504	504
<i>p-value</i> Random match = Ability match & instructions	0.155	0.597	0.254	0.372	0.017	0.455

Note: Given the outcomes are measured on a five-point Likert scale, estimates are derived from ordered logistic regressions; Clustered standard errors 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; *p-value* at the bottom is derived from the *F*-test comparing the estimate 'Treat' (*Random match*) with the combined estimate 'Treat + Treat * matching + Treat * instruction' (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation.

Table B.V: Heterogeneous Effects by SES and Gender

Dependent variable: Course Financial Proficiency	SES		Difference <i>p-value</i>	Gender		Difference <i>p-value</i>
	<i>Low</i>	<i>High</i>		<i>Male</i>	<i>Female</i>	
Treat	0.108 (0.102)	0.198** (0.0773)	0.394	0.128 (0.100)	0.228*** (0.0798)	0.369
Treat * matching	0.162 (0.101)	0.0926 (0.0880)	0.534	0.117 (0.105)	0.0952 (0.0895)	0.860
Treat * instruction	-0.0572 (0.0972)	-0.0341 (0.0949)	0.839	-0.0834 (0.106)	0.0503 (0.108)	0.352
<i>Controls</i>	Yes	Yes		Yes	Yes	
<i>Obs.</i>	798	1,609		1,119	1,288	
<i>p-value</i> Random match = Ability match & instructions	0.347	0.478		0.737	0.123	

Note: Clustered standard errors at class level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Controls: pre-treatment value of outcome, (gender for SES subgroup analysis), (SES for gender subgroup analysis), track, type of school, grade; *p-value* at the bottom is derived from the *F*-test comparing the estimate 'Treat' (*Random match*) with the combined estimate 'Treat + Treat * matching + Treat * instruction' (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation; Low SES students are defined as travelling once or zero times, high SES students as travelling two or more times.

Table B.VI: Robustness Test – School Level Clustered Standard Errors

Dependent variable	Course Financial Proficiency		Financial Knowledge		Financial Behaviour	
Treat	0.227** (0.0923)	0.177* (0.0915)	0.254** (0.101)	0.206** (0.100)	0.115** (0.0550)	0.0671 (0.0547)
Treat * matching	0.0780 (0.0917)	0.104 (0.0804)	0.0303 (0.105)	0.0604 (0.100)	0.105 (0.0680)	0.122** (0.0571)
Treat * instruction	-0.179* (0.107)	-0.0115 (0.0799)	-0.177 (0.112)	-0.0407 (0.0875)	-0.129 (0.0946)	0.0357 (0.0867)
<i>Controls</i>	No	Yes	No	Yes	No	Yes
<i>Obs.</i>	2,407	2,407	2,407	2,407	2,407	2,407
<i>p-value</i> Random match = Ability match & instructions	0.287	0.271	0.146	0.825	0.761	0.048

Note: Clustered standard errors at school 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; *p-value* at the bottom is derived from the *F*-test comparing the estimate ‘Treat’ (*Random match*) with the combined estimate ‘Treat + Treat * matching + Treat * instruction’ (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation.

Table B.VII: Robustness Test – Instrumental Variables

Dependent variable	Course Financial Proficiency		Financial Knowledge		Financial Behaviour	
Treat	0.290** (0.127)	0.279** (0.127)	0.301* (0.175)	0.293 (0.182)	0.142 (0.105)	0.128 (0.0902)
Treat * matching	0.236 (0.183)	0.166 (0.160)	0.210 (0.206)	0.139 (0.195)	0.183 (0.184)	0.132 (0.158)
Treat * instruction	-0.348** (0.162)	-0.120 (0.126)	-0.333** (0.144)	-0.140 (0.117)	-0.230 (0.176)	-0.0175 (0.150)
<i>Controls</i>	No	Yes	No	Yes	No	Yes
<i>Obs.</i>	1,516	1,516	1,516	1,516	1,516	1,516
<i>p-value</i> Random match = Ability match & instructions	0.414	0.728	0.496	0.993	0.698	0.305

Note: Clustered standard errors 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; *p-value* at the bottom is derived from the *F*-test comparing the estimate ‘Treat’ (*Random match*) with the combined estimate ‘Treat + Treat * matching + Treat * instruction’ (*Ability match & instruction*) in order to examine the value added of the combination of grouping and instruction as compared to no differentiation.

Table B.VIII: Robustness Test – Coarsened Exact Matching

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

Note: Estimates are derived from three separate regressions including binary treatment variables; Clustered standard errors at class level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Controls for ITT: pre-treatment value of outcome, gender, SES, track, type of school, grade; Controls 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, language, pre-treatment value of outcome, importance of financial literacy, and grade.

Appendix C: 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

Appendix D: Attrition Analysis for Second Post-treatment Financial Proficiency Test

From the 1,557 students in all three experimental conditions who took the first post-treatment financial literacy test, 1,199 (77 percent) did not complete the second post-treatment test (885 (76.49 percent) in study I and 314 (78.5 percent) in study II). Given the large attrition rate, we test for selective attrition across experimental conditions and find that students who were matched based on their ability, receiving uniform or differentiated instructions, during the lectures were significantly more likely to complete the homework, as presented in Table D.I.

Given the pre- and first wave of post-treatment data were collected for all students, we compare non-complying (only completing the first post-test) and complying students (completing the second post-test as well) within each condition. Table D.II reveals significant differences for various student characteristics. In particular, for the condition where students are matched according to their ability and additional instructions are provided to lower ability pairs, we observe that the majority of the students completing the homework attended an academic track and performed significantly better on the first post-treatment test as compared to those not completing the homework. Accordingly, we must interpret the results on the second post-treatment test with caution.

Table D.I: Selective Attrition in Second Post-test

Dependent variable	Attrition Study I & II
Matching	-0.130* (0.0724)
Instruction	-0.135 (0.0951)
<i>Controls</i>	Yes
<i>Obs.</i>	1,557
<i>p-value</i> Random match = Ability match & instructions	0.005

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; *p-value* at the bottom is derived from the *F*-test testing 'Matching + Instruction = 0' (*Ability match & instruction*).

Due to the limited sample for randomly-paired students in the second post-treatment financial literacy test (55 observations), we are unable to explore potential heterogeneous effects.

Table D.II: Comparison of Non-Complying and Complying Students Within Conditions

STUDY I	Variables	Random match	<i>p</i> -value	Ability match	<i>p</i> -value	Ability match & instruction	<i>p</i> -value
<i>Percentage of Attrition</i>		93.00%		79.37%		57.84%	
<i>Number of Attrited Students</i>		372		277		236	
Type of school (private)		0.171 (0.159)	0.282	-0.101 (0.061)	0.109	0.009 (0.089)	0.922
Class size		2.628 (3.723)	0.485	-1.739 (1.057)	0.111	1.253 (1.486)	0.405
Track	<i>Academic</i>	31.49%		-15.52%		9.08%	
	<i>Technical</i>	-25.96%	0.273	5.05%	0.045	-10.00%	0.716
	<i>Vocational</i>	-5.53%		10.47%		0.92%	
<i>Background characteristics</i>							
Gender (female)		0.209 (0.075)	0.009	-0.340 (0.159)	0.042	0.004 (0.072)	0.953
Age (years)		-0.038 (0.109)	0.730	0.043 (0.055)	0.440	-0.168 (0.091)	0.072
Number of holidays per year (4)		-0.030 (0.248)	0.905	-0.077 (0.156)	0.623	0.133 (0.139)	0.345
Language (Dutch)		-0.044 (0.070)	0.535	0.003 (0.040)	0.933	0.217 (0.120)	0.078
Grade in mathematics (5)		-0.416 (0.129)	0.003	-0.083 (0.131)	0.534	0.367 (0.228)	0.116
<i>Financial characteristics</i>							
Importance of financial literacy (5)		-0.193 (0.178)	0.286	-0.011 (0.061)	0.857	-0.068 (0.116)	0.565
Importance of saving (5)		-0.325 (0.114)	0.007	-0.092 (0.125)	0.469	-0.010 (0.134)	0.939
Comparison shops before purchase (5)		0.567 (0.245)	0.027	0.054 (0.117)	0.647	-0.187 (0.151)	0.221
<i>Pre-treatment financial scores</i>							
Course financial proficiency (9)		0.200 (0.175)	0.260	-0.573 (0.250)	0.030	0.119 (0.285)	0.679
Financial knowledge (6)		0.367 (0.120)	0.004	-0.433 (0.186)	0.028	0.038 (0.196)	0.848
Financial behaviour (3)		-0.167 (0.075)	0.034	-0.141 (0.100)	0.173	0.081 (0.135)	0.552
<i>1st post-treatment financial scores</i>							
Post Course financial proficiency (9)		0.388 (0.200)	0.061	-0.036 (0.355)	0.920	0.101 (0.248)	0.685
Post Financial knowledge (6)		0.417 (0.168)	0.018	-0.027 (0.229)	0.906	0.017 (0.159)	0.915
Post Financial behaviour (3)		-0.029 (0.123)	0.814	-0.009 (0.149)	0.954	0.084 (0.125)	0.506
STUDY II	Variables	Random match	<i>p</i> -value	Ability match	<i>p</i> -value	Ability match & instruction	<i>p</i> -value
<i>Percentage of attrition</i>		83.23%		75.68%		74.73%	
<i>Number of Attrited Students</i>		134		112		68	
Type of school (private)		0.296 (0.202)	0.159	-0.107 (0.081)	0.213	0.056 (0.276)	0.844
Class size		6.082 (1.911)	0.005	2.116 (3.051)	0.501	0.407 (2.776)	0.886
Track	<i>Academic</i>	44.44%		17.56%		-65.22%	
	<i>Technical</i>	-25.93%	0.096	2.18%	0.472	-18.60%	0.001
	<i>Vocational</i>	-18.51%		-19.74%		83.82%	
<i>Background characteristics</i>							
Gender (female)		0.138 (0.066)	0.051	-0.092 (0.153)	0.558	0.110 (0.165)	0.520
Age (years)		-0.051 (0.081)	0.536	-0.129 (0.266)	0.636	0.283 (0.179)	0.143
Number of holidays per year (4)		0.408 (0.232)	0.095	0.286 (0.319)	0.388	-0.354 (0.187)	0.085
Language (Dutch)		-0.119 (0.030)	0.001	0.171 (0.117)	0.171	0.054 (0.063)	0.414
Grade in mathematics (5)		0.600 (0.433)	0.181	0.108 (0.190)	0.580	0.644 (0.360)	0.101
<i>Financial characteristics</i>							
Importance of financial literacy (5)		0.225 (0.124)	0.087	0.090 (0.231)	0.703	-0.115 (0.164)	0.497
Importance of saving (5)		0.202 (0.304)	0.515	-0.284 (0.154)	0.091	-0.670 (0.111)	0.000
Comparison shops before purchase (5)		0.198 (0.164)	0.243	-0.135 (0.225)	0.559	-0.270 (0.228)	0.261
<i>Pre-treatment financial scores</i>							
Course financial proficiency (9)		0.608 (0.375)	0.121	-0.228 (0.630)	0.724	-1.869 (0.288)	0.000
Financial knowledge (6)		0.530 (0.280)	0.074	0.002 (0.536)	0.997	-1.070 (0.211)	0.000
Financial behaviour (3)		0.078 (0.136)	0.572	-0.230 (0.146)	0.142	-0.800 (0.147)	0.000
<i>1st post-treatment financial scores</i>							
Post Course financial proficiency (9)		-0.311 (0.500)	0.540	-0.000 (0.717)	1.000	-2.136 (0.524)	0.002
Post Financial knowledge (6)		-0.277 (0.316)	0.391	0.179 (0.439)	0.691	-0.930 (0.474)	0.075
Post Financial behaviour (3)		-0.034 (0.252)	0.894	-0.179 (0.315)	0.581	-1.206 (0.203)	0.000

Note: Differences in means between attrited and non-attrited students (completing the second 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 track, relative differences are given.