

This item is the archived peer-reviewed author-version of:

The effects of computer-assisted adaptive instruction and elaborated feedback on learning outcomes : a randomized control trial

Reference:

Itebeke Kaat, De Witte Kristof, Schelfhout Wouter.- The effects of computer-assisted adaptive instruction and elaborated feedback on learning outcomes : a randomized control trial
Computers in human behavior - ISSN 0747-5632 - 120(2021), 106666
Full text (Publisher's DOI): <https://doi.org/10.1016/J.CHB.2020.106666>
To cite this reference: <https://hdl.handle.net/10067/1763430151162165141>

The effects of computer-assisted adaptive instruction and elaborated feedback on learning outcomes. A randomized control trial. †

Kaat Iterbeke^{*,a}, Kristof De Witte^{a,b}, and Wouter Schelfhout^c

^a *Leuven Economics of Education Research, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium*

^b *UNU-MERIT, Maastricht University, Boschstraat 24, 6211 AX Maastricht, the Netherlands*

^c *Antwerp School of Education, Universiteit Antwerpen, Venusstraat 35, 2000 Antwerpen, Belgium*

Abstract. Using a computer-based learning environment, the present paper studied the effects of adaptive instruction and elaborated feedback on the learning outcomes of secondary school students in a financial education program. We randomly assigned schools to four conditions on a crossing of two factors: the type of instruction (uniform or adaptive) and feedback (verification or elaborated). A total of 1,177 students in 32 schools completed the program in ability groups in the classroom. The results showed that the program, on average, enhanced the financial knowledge of students by almost half of a standard deviation. No significant changes in students' financial behavior were found. Despite the promise of adaptive practices to address the individual needs of students, we observed no additional learning gains associated with adaptive instruction and elaborated feedback. A marginally significant heterogeneous effect for gender was reported, where girls were negatively affected by adaptive instruction. Moreover, despite our sample included more students from a favorable socioeconomic status, the adaptive practices seemed to lower the motivation level. Hence, while no information on the time spent on the instruction and feedback was retrieved, the latter finding suggested that the practices may have been perceived as burdensome by students, thereby rendering them ineffective.

Keywords. Computer-Assisted Learning; Adaptive Instruction; Feedback; Financial Literacy; Randomized Control Trial.

† This work was supported by the Flemish Science Organisation [grant S000617N]. This work is registered in the AEA RCT Registry and the unique identifying number is AEARCTR-0004431. We are grateful to Boukje Compen, Joana Elisa Maldonado, Koen Declercq, Lieve Lammens, Johan Mestdagh, Els Lagrou, and Danièle Van der Espt for their research assistance. We thank seminar participants in Maastricht and Leuven for valuable suggestions on an earlier version of this paper. Declarations of interest: none.

* Corresponding author.

E-mail addresses: kaat.iterbeke@kuleuven.be (K. Iterbeke), kristof.dewitte@kuleuven.be, k.dewitte@maastrichtuniversity.nl (K. De Witte), wouter.schelfhout@uantwerpen.be (W. Schelfhout).

1 Introduction

Academic diversity is omnipresent in today's classrooms. As differences between students matter greatly in learning, practices that address the individual needs of students are key to ensure all students receive appropriate academic challenges (Tomlinson et al., 2003). The supremacy of personalized learning over one-size-fits-all approaches is well-established in the literature (for review see e.g., Deunk et al. (2015) and Kulik et al. (1990)). Yet, actual implementation in school lags behind due to time, knowledge, and practical constraints present in traditional teaching (Aleven et al., 2017). Consequently, the use of technology in education increased in recent years, offering the possibility of personalized learning (Deunk et al., 2015). Meta-analyses on computer-based learning environments indicate positive effects compared with traditional teaching methods (Cheung & Slavin, 2013; Tamin et al., 2011), in particular when the environment is adaptive to the needs of students (Aleven et al., 2017; Faber et al., 2017). However, despite the large amount of research on the effectiveness of technology as a direct input in the education production function, it remains unclear, when, how, and which particular components of the computer-based learning environment contribute to learning (Shute & Zapata-Rivera, 2012; Faber et al., 2017). For instance, Muralidharan et al. (2019) provide a review of empirical evidence on computer-assisted instruction. Despite large heterogeneity in effects, they observed that interventions delivering the largest effects incorporated an adaptive component (e.g., Banerjee et al., 2007). However, as the earlier literature focused on the value-added of computer-assisted instruction, no distinction was made in the various components of the instruction (Muralidharan et al., 2019). The present paper contributes to the literature by examining the combined and isolated effects of two commonly-used components, i.e., adaptive instruction and feedback, in a computer-assisted program.

To establish our results, we conducted a randomized control trial in the eighth and ninth grade of Flemish secondary education involving 1,177 students in 32 schools. The computer-assisted program dealt with financial literacy education. Financial education is an attractive subject for our study as large discrepancies in financial literacy among youth are found. While family, sociodemographic, and peer characteristics can explain the differences in financial literacy, cognitive ability also appears to be an important determinant, i.e., financial knowledge is substantially higher for high-ability students (Lusardi et al., 2010). Given these differences, one may question the effectiveness of financial education when all students are taught the same program uniformly.¹ As suggested by Kaiser and Menkhoff (2019), a differentiated approach may be more effective.

2 Theoretical framework and literature review

Adaptivity in computer-based learning environments has been reviewed thoroughly in the literature (e.g., Aleven et al., 2017; Shute & Zapata-Rivera, 2012; Vandewaetere et al., 2011) and builds on multiple theoretical perspectives, including the zone of proximal development (ZPD), the model-

¹ Experimental research on school-based financial education programs includes, among others, Bruhn et al., 2016; Bover et al., 2018; Compen et al., 2020; Frisnacho, 2018; Hinojosa et al., 2010; Lührmann et al., 2018; Maldonado et al., 2019.

scaffold-fade paradigm, and the assistance dilemma (Aleven et al., 2017). The present study investigates the importance of tailoring instructions and feedback to the individual learning needs and abilities of students. In particular, our approach of assistance is to offer personalized learning material, varying in task difficulty and amount of instruction, after an assessment of students' prior knowledge about the topic. In addition, the learning material includes feedback with the aim of reducing the gap between current understandings and the learning goal (Hattie & Timperley, 2007). Resolving this gap can, for instance, reduce uncertainty about how well (or poorly) the student is performing and provide information that helps correct erroneous learning strategies (Shute, 2008).

By offering the adaptive practices in this manner, we build on the ZPD theory, as further applied in computer-based learning environments with adaptive instruction (Sottolare & Goldberg, 2012; Subban, 2006) and feedback (Golke et al., 2015). According to Vygotsky's theory (1987), new learning occurs when students are offered a moderate challenge, i.e., when the learning material is adjusted to their appropriate *learning zone*. The zone refers to a required level of mastery where students cannot learn independently but can succeed with scaffolding (Tomlinson et al., 2003). A way of accessing students' zone of proximal development and hence, avoiding intrinsic cognitive overload (difficulty associated with the learning content) due to a challenge level that is too high, is via adaptive practices (Sottolare & Goldberg, 2012; Subban, 2006). In a similar vein, Hollender et al. (2010) state that the objective of adaptive practices should be to adapt students' intrinsic load, while simultaneously reducing the extraneous cognitive load (the way information is presented to students).

2.1 Adaptive instruction

Scholars underlined the importance of adequately identifying student characteristics – such as prior knowledge, cognitive ability, interests – and developed multiple approaches for adaptivity, varying in granularity (Vandewaetere et al., 2011).

More fine-grained ways of adaptive instruction, such as Intelligent Tutoring Systems, take into account multiple student characteristics and provide moment-by-moment adjustment of instruction. Despite their effectiveness in terms of enhancing student learning (Kulik & Fletcher, 2016), Shute and Zapata-Rivera (2012) argue that these complex techniques also come at a cost.

Other types of adaptive instruction, on the other hand, lack univocal empirical verification. For instance, using a large-scale randomized control trial in Dutch secondary schools, van Klaveren et al. (2017) compared the effectiveness of static and adaptive practice programs. While no significant effect of the adaptive version was found for the average student, high-ability students achieved lower scores when practicing adaptively. Per contra, also in the Netherlands, but on a smaller scale, Haelermans et al. (2015) and Haelermans and Ghysels (2017) found positive effects of adaptive instruction on students' numeracy performance and performance in a biology class. Iterbeke et al. (2020) examined the impact of ability matching and a basic (non-personalized) form of adaptive instruction in a financial education program in Flanders. Although student learning outcomes, on average, were not altered by the practices,

non-native students significantly benefited from ability matching, conditional on receiving adaptive instruction.

The effects found in aforementioned studies seem to depend on particular student characteristics. This is in line with research on standard computer-assisted programs indicating heterogeneity by multiple student and classroom characteristics, such as the prior knowledge, class size, and class heterogeneity (Barrow et al., 2009).

2.2 Feedback

Despite the existence of many models and methods for providing feedback (Mory, 2004), research has shown that they are not equally effective in terms of enhancing student performance (Hattie, 2009). Simply adding feedback does not guarantee that students seek and process the feedback, and many factors appear to influence students' willingness to engage with it (Maier et al., 2016; Timmers et al., 2013). Not only does this depend on the content of feedback, the timing, and the context in which feedback is given (Hattie & Timperley, 2007), student characteristics (e.g., prior knowledge, gender, motivational state) also appear to moderate the effect of feedback (e.g., Fyfe & Rittle-Johnson, 2016; Narciss et al., 2014; Timmers & Veldkamp, 2011).

In this study, we focus on two common types of feedback, i.e., verification feedback, where a simple judgment is made of whether an answer is correct, and elaborated feedback, where information on the (in)correctness of an answer and relevant cues to guide the student towards the correct answer are provided (Shute, 2008). It is unclear today whether elaborated feedback outperforms verification feedback. While a meta-analysis by van der Kleij et al. (2015) demonstrated that elaborated feedback is more effective than verification feedback when students engage in complex tasks, recent studies reported no (or limited) additional effects of elaborated feedback (Attali & van der Kleij, 2017; Fyfe, 2016; Golke et al., 2015). Although more specific feedback is generally more effective, Shute (2008) argues that, if feedback is too lengthy or complex, it may lead to a cognitive burden in students, rendering it ineffective. Further, it is argued that particular features in the learning environment, the learning domain, and learner characteristics influence the effects of feedback complexity (Wang et al., 2019).

2.3 Research questions and hypotheses

The financial education literature shows that school-based financial education programs improve the financial knowledge and, to a smaller extent, the financial behavior of students (Kaiser & Menkhoff, 2019). Yet, given the discrepancies in financial literacy among youth, it is suggested that a differentiated approach is more effective. In the present paper, we explore the importance of tailoring a computer-assisted financial education program to the needs of students via adaptive instruction and elaborated feedback. In particular, we are interested in the learning gains associated with adaptive instruction compared to one-size-fits-all instruction and the learning gains associated with adaptive instruction combined with elaborated feedback compared to adaptive instruction combined with verification

feedback. As previous research indicates that the effects of the two adaptive practices are ambiguous and seem to depend on various factors (e.g., Barrow et al., 2009; Wang et al., 2019), we also examine the importance of multiple student and class characteristics. The aims of our study are addressed with the following questions and subsequent research hypotheses.

RQ1. What is the impact of a financial education program on the financial proficiency of secondary school students?

H1. Students' financial knowledge will improve after the program. A smaller improvement in the financial behavior of students is expected.

RQ2. What are the learning gains associated with an adaptive version of the financial education program? In particular, what are the effects of adaptive instruction and elaborated feedback in the program?

H2a. An adaptive version of the financial education program is expected to enhance the financial outcomes of students on average.

H2b. The effects of adaptive instruction and elaborated feedback are a priori ambiguous.

RQ3. Are the effects of adaptive instruction and elaborated feedback heterogeneous with respect to multiple student and class characteristics?

H3. The effectiveness of adaptive instruction and elaborated feedback will be moderated by multiple student and class characteristics (e.g., prior knowledge, motivational state, class size).

3 Method

3.1 Design

Randomization was conducted at the school level in order to avoid contamination effects. To improve power in the relatively small sample of schools (Bruhn & McKenzie, 2009) and to ensure balance across conditions *ex ante*, we performed a stratified randomization procedure at school level via a computer algorithm. In particular, schools were first partitioned into strata according to the following characteristics, the educational network (private or public), the participating grades in the school (eighth grade, ninth grade, or both), the school size (the number of students the school was participating with), and the education tracks offered (academic, technical/vocational, or both). Then, within each stratum, schools were randomized to a condition.

Table I: *Control and Experimental Conditions*

	Control condition	Experimental condition I	Experimental condition II	Experimental condition III
		Uniform, EF	Adaptive, EF	Adaptive, VF
Learning path	No	Intermediate	Adaptive	Adaptive
Feedback	No	Elaborated	Elaborated	Verification

Note: EF refers to elaborated feedback; VF refers to verification feedback.

In order to examine the program effectiveness and (un)combined impact of adaptive instruction and elaborated feedback, we designed four conditions, i.e., one control condition and three experimental

conditions. The control condition did not receive the financial education program. In the first experimental condition, all students followed an intermediate-level learning path, regardless of their proficiency level. Elaborated feedback was given after all exercises. In the second experimental condition, students followed an adaptive learning path with elaborated feedback. Note that a comparison with the first experimental condition allowed us to evaluate the impact of tailoring instructions to students' needs. Finally, in the third experimental condition, in order to study the importance of elaborated feedback, students followed an adaptive learning path, however, they were only given verification feedback.² Table I provides a summary of the conditions.

3.2 Participants

Teachers from 46 schools with 180 eighth and ninth grade classes signed up to participate in the financial education program in Flanders, the northern region of Belgium. Details on the Flemish context and curriculum are provided in Appendix A. Although 1,921 students took the pre-treatment test, data from 744 students were excluded from the analysis as they did not take the post-treatment test. The final sample included 1,177 students in 94 classes in 32 schools. The background characteristics of the students are summarized in section 4.

3.3 Material and measures

3.3.1 Pre-treatment test

Prior to the lectures, all students took a test to measure their baseline financial proficiency. The pre-treatment test was designed as a computer-aided multiple-choice test. The test included several questions on students' demographics and eight questions that referred directly to the material and measured students' financial proficiency. Financial proficiency can be decomposed into financial knowledge and financial behavior. We assessed financial knowledge by five questions covering the calculation of monthly savings, risks and rewards related to investment products, rates of return, interest, and inflation (similar to questions 1, 2, 3, 4, and 6 in Appendix D). Two questions for the latter three financial concepts were taken from Lusardi and Mitchell (2011). In line with the content of the lectures, we measured financial behavior by three questions related to the reliability of information and saving strategies (similar to questions 5, 7, and 8 in Appendix D).

3.3.2 Learning material

The learning material was designed as four lectures of 50 minutes in the form of a computer-assisted learning path. The learning path focused on the topics of saving and investing, which are part of the content areas of the PISA financial literacy assessment, *Planning and Managing Finances* and *Risk and Reward* (OECD, 2016). After the lectures, students were expected to know how to draw up a budget to

² We rule out the presence of confounding variables by comparing the program effectiveness for average-performing students following the intermediate-level learning path in either the *uniform, elaborated feedback* condition or the *adaptive, elaborated feedback* condition. Given both experimental conditions *by definition* are equal for those students (i.e., intermediate-level learning path with EF), we expect those students to perform equally well in the post-test. This is confirmed in Appendix F.I.

plan spending and saving, to understand the benefits of saving for long-term goals or anticipated changes, to recognize certain financial products and investments, and to know about risks and rewards associated with substitutes for financial products.

The path consisted of six modules and each module had three independent parts, i.e., information sheets, multiple exercises, and a formative test. The exercises comprised, among others, multiple-choice quizzes, fill-in-the-blank and drag-and-drop exercises, learning games, interactive videos, and case studies. For fill-in-the-blank and drag-and-drop exercises, students were requested to repeat the exercise until they answered correctly. The formative tests informed students on the learning goals of each module and the extent to which they had reached these goals.

3.3.3 *First post-treatment test*

To measure the short-term impact of the financial education program, a post-treatment computer-aided multiple-choice test was administered at the start of the last lecture for students in the experimental conditions.³ Students in the control condition completed the test during the same period. The financial proficiency questions were constructed by rephrasing and using adjustment of numbers of the pre-test questions.⁴ In addition to these questions, the test included four items related to students' motivational state during the lectures, which were based on the *Motivation Strategies for Learning Questionnaire* (Pintrich et al., 1993). In particular, the items evaluated how interesting, important, and useful students perceived the lectures and the learning material. Responses were given on a seven-point Likert-scale ranging from 'totally disagree' (1) to 'totally agree' (7). Based on these items, we constructed an overall motivation measure ($\alpha = 0.83$).

3.3.4 *Second post-treatment test*

To measure the long-term impact, a second post-treatment test was offered as a computer-aided homework assignment for students in the experimental conditions and included similar financial proficiency questions as the other two tests.

3.4 Procedure

To ensure a uniform program implementation, teachers were requested to provide the program during pre-specified periods and regular class hours. By means of the computer-based learning environment, which automatically provided both instruction and feedback, teachers were expected to intervene as little as possible, i.e., they were only asked to give a brief whole-class introduction on the purpose of

³ As some students worked at a slower pace than others, we required teachers to plan the test at the start of the last lecture to guarantee all students in the classroom were able to take the full test. After completion of the test, students were requested to resume the learning path.

⁴ The pre- and post-treatment questions on financial proficiency were face validated by two senior teachers in the research team and showed good construct validity. As each question measured very specific financial knowledge or behavior, the internal consistency was low ($\alpha = 0.43$ and 0.56 for the pre- and post-treatment questions, respectively). All post-treatment questions showed fair or good discrimination (Pearson Product Moment correlation of 0.10 or higher) and difficulty indices ranged from 0.31 to 0.76 .

the lectures and instructions for the learning path. Besides a manual, no specific teacher professional development was provided. Accordingly, we aimed to minimize teacher impact to prevent potentially confounding influences from affecting the program effectiveness.

Students followed the learning path in pairs of two. Student pairs were formed by the teacher before the start of the lectures using students' grades in mathematics. This ensured that students were ability-grouped.

According to the experimental condition the school was allocated to, student pairs either followed an intermediate-level learning path or a path adapted to their financial proficiency. In order to determine their proficiency level, student pairs were requested to complete a short diagnostic test at the start of the path. When an adaptive learning path was followed in the school, student pairs were assigned to one out of three paths based on their proficiency level. Students were unaware of the level of each path and the paths had identical learning objectives. The paths were tailored to the needs of student pairs such that the basic path, aimed for students with low baseline financial proficiency, consisted of basic exercises and language, hints and cues, and additional explanation. The intermediate-level path included intermediate-level exercises and several hints. Finally, the advanced-level path did not comprise hints or cues and students were challenged with more difficult exercises.

Table II: Differences in Feedback Type

Feedback type	Type of exercise	Verification feedback	Elaborated feedback
Knowledge of result (right-wrong)	<i>All exercises & formative tests</i>	X	X
Correct response	<i>Multiple-choice, formative tests</i>		X
Error flagging & try again	<i>Drag-and-drop, fill-in-the-blanks</i>		X
Hints & cues	<i>Multiple-choice</i>		X
Information on misconceptions	<i>Multiple-choice, formative tests</i>		X
Reteaching of material	<i>Multiple-choice</i>		X
Additional exercises	<i>Multiple-choice</i>		X

Note: The type of elaborated feedback for multiple-choice quizzes varied depending on the module; Error flagging highlights errors in a solution, without giving the correct answer.

The learning path provided immediate item-specific feedback, either in a basic (verification feedback) or extended (elaborated feedback) format, depending on the experimental condition. Verification feedback informed students about the correctness of the answer, however, it did not provide the correct answer or additional information. Depending on the type of exercise, elaborated feedback, on the other hand, took many forms, i.e., information on why the incorrect answer was wrong and why the correct answer was right, error flagging and try again, strategic hints, worked examples, additional exercises, reteaching of particular content, and additional information on the topic in case students answered incorrectly. Table II indicates the differences in the type of feedback used in our study and under what conditions (type of exercise) which type of elaborated feedback was provided. Figure B.I provides an example of the differences in the adaptive learning path, while Figure B.II provides an example of the feedback in the learning path.

The pre- and first post-treatment tests took 15-20 minutes, on average, and were administered during regular class hours under the supervision of the teacher. Students in control schools completed the tests at the same time as students in treatment schools. Teachers in the control schools obtained the material after their students took the pre- and first post-treatment test as an incentive to comply with the prescribed instructions. Students completed the tests individually. They were unaware of how well they performed on the tests and the scores were not part of the normal course assessment. Approximately four weeks after the lectures, students in the treatment schools completed the second post-treatment test as a homework assignment. Figure I presents the program implementation for the control and experimental conditions.

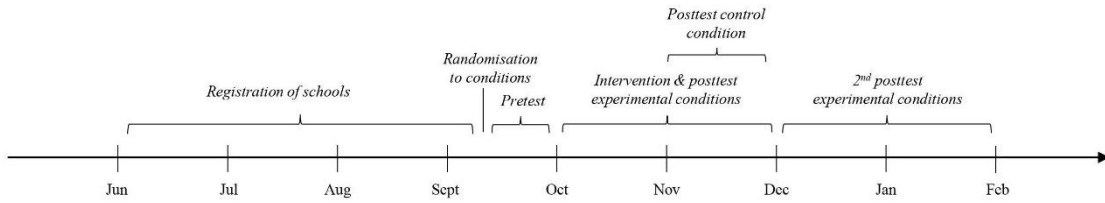


Figure I: Timeline of the Program Implementation for the Control and Experimental Conditions

3.5 Empirical strategy

To estimate the causal impact of the treatments, we analyzed the following Intent-to-Treat regression model:

$$y_{is}^1 = \alpha + \beta_0 Treat_{is} + \beta_1 adaptive_{is} + \beta_2 EF_{is} + \beta_3 y_{is}^0 + \Sigma \beta_4' X_i + \Sigma \beta_5' S_{is} + \epsilon_{is} \quad (1)$$

Let y_{is}^1 be the standardized value of an outcome measure, i.e., financial proficiency, financial knowledge, or financial behavior for student i in school s . $Treat_{is}$ takes value one if student i in school s received the financial education program, zero otherwise. $adaptive_{is}$ takes value one if student i in school s received the adaptive learning path, zero otherwise. EF_{is} takes value one if student i in school s received elaborated feedback in the learning path, zero otherwise. By experimental design $Treat_{is}$ takes value one whenever $adaptive_{is}$ takes value one and/or EF_{is} takes value one. Accordingly, $adaptive_{is}$ captures the average incremental gain in an outcome measure for a student receiving the adaptive learning path as compared to a uniform learning path. EF_{is} then captures the average incremental gain in an outcome measure for a student receiving elaborated feedback in the adaptive learning path as compared to verification feedback in the adaptive learning path.

To this basic specification, we added student characteristics X_i in a second specification to improve the precision and to account for possible baseline differences. Moreover, we added y_{is}^0 , the pre-treatment value of an outcome, for student i in school s and a time indicator between pre- and post-treatment tests for student i .⁵

⁵ As teachers had the liberty to plan the lectures in a relatively long period after the pre-treatment test, some students completed the post-treatment test shortly after the pre-test, whereas others, for example, after four weeks. Hence, we control for the time between the tests.

Finally, following Bruhn and McKenzie (2009), a third specification included strata fixed effects S_{is} to account for the fact that the randomization was done after stratification based on several school characteristics, i.e., the educational network, grades, school size, and education tracks. The strata variable included 19 values corresponding to the 19 different strata in our final sample.

Given randomization occurred at the level of the school, all specifications included standard errors clustered at the school level s to take into account the multi-level data structure and to allow for within-cluster dependence. However, a practical limitation of inference with cluster-robust standard errors is that its asymptotic justification assumes the number of clusters to go to infinity. With the few (treated) clusters in our data, cluster-robust standard errors were likely to be downward biased. We solved for potential bias via a wild cluster restricted bootstrap approach proposed by Cameron et al. (2008). They suggest to cluster the standard errors, however, apply bootstrapping to obtain bootstrap critical values, such that an asymptotic refinement is provided for the few clusters. Mackinnon and Webb (2018) show that wild cluster restricted bootstrap (with the null hypothesis imposed) generates more conservative p -values than the unrestricted approach and that it tends to moderately under-reject. Hence, we considered this to be the most conservative approach in our setting.⁶

4 Data

Table III presents the school and student characteristics for the final sample involving 1,177 students, 94 classes, and 32 schools which were randomly assigned to one of the conditions. The number of students and schools are not equal across conditions, which can partially be explained by attrition (see section E.1 in Appendix E). Overall, the characteristics are relatively balanced across conditions.

The first panel indicates the sample to include mostly private schools, which is in line with the overall majority of private schools in the Flemish education system. On average, teachers taught in a class of 17 students. The majority of students participating in the program were in the eighth grade.

Student background characteristics, as presented in panel B, show that, on average, 50 percent of students were female, 88 percent spoke Dutch (the official language) at home, and students were 13 and half years old, on average. Most students found financial literacy important (mean Likert scale of 4.16 out of 5). Both for the education track and grade in mathematics, we find the *adaptive, elaborated feedback* condition to have lower-ability students compared to the control condition. It should be noted that students' math proficiency was approximated by students' self-assessed grade in mathematics on a five-point scale and that it greatly depended on their school and education track (as there are no national standardized examinations in Flanders). To account for the imbalances, we control for these variables in the regression analyses.

⁶ See Mackinnon and Webb (2018) for alternative approaches to solve the problem of few clusters. Note that, as the restricted and unrestricted variants of the wild cluster bootstrap yield similar inferences in our case, there is no real need to use other approaches.

Table III: Descriptive Statistics

Variables	Control	Uniform, EF	<i>p</i> -value	Adaptive, EF	<i>p</i> -value	Adaptive, VF	<i>p</i> -value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. School characteristics							
Private education	0.84	0.83	0.955	0.97	0.277	0.85	0.960
Class size	19.11 (6.16)	18.12 (4.14)	0.585	15.46 (5.01)	0.137	20.89 (4.06)	0.324
Fraction of 8 th grade students	0.64	0.72	0.796	0.57	0.740	0.78	0.471
Panel B. Background characteristics							
Track							
Academic	295 (94.55)	135 (84.91)		217 (51.06)		264 (93.95)	
Technical	17 (5.45)	21 (13.21)	0.489	146 (34.35)	0.034	10 (3.56)	0.646
Vocational	0 (0.00)	3 (1.89)		62 (14.59)		7 (2.49)	
Gender (female)	0.57	0.50	0.546	0.44	0.229	0.50	0.386
Age (years)	13.42 (0.65)	13.35 (0.60)	0.852	13.73 (0.88)	0.262	13.35 (0.66)	0.768
Language (Dutch)	0.91	0.90	0.796	0.84	0.228	0.85	0.274
Grade in mathematics (5-points)	3.69 (1.03)	3.71 (1.03)	0.925	3.25 (1.15)	0.074	3.38 (1.11)	0.219
Importance financial literacy (5-points)	4.22 (0.74)	4.15 (0.62)	0.572	4.11 (0.75)	0.363	4.14 (0.68)	0.534
Panel C. Pre financial scores							
Financial proficiency (8-points)	4.03 (1.56)	3.99 (1.60)	0.919	3.55 (1.72)	0.137	3.52 (1.73)	0.147
Financial knowledge (5-points)	2.18 (1.09)	2.26 (1.03)	0.641	2.04 (1.13)	0.405	2.07 (1.08)	0.567
Financial behavior (3-points)	1.85 (0.93)	1.73 (0.95)	0.508	1.51 (0.99)	0.041	1.44 (1.02)	0.036
Panel D. Post financial scores							
Financial proficiency (8-points)	4.40 (1.73)	5.40 (1.48)	0.049	4.34 (2.03)	0.868	4.57 (1.95)	0.508
Financial knowledge (5-points)	2.65 (1.19)	3.42 (1.08)	0.007	2.78 (1.37)	0.580	2.91 (1.28)	0.054
Financial behavior (3-points)	1.75 (0.91)	1.99 (0.84)	0.323	1.56 (0.98)	0.269	1.65 (0.94)	0.429
<i>Number of Students</i>	312	159		425		281	
<i>Number of Schools</i>	9	5		10		8	

Note: EF refers to elaborated feedback, VF refers to verification feedback; Mean value and standard deviation in parentheses; Absolute and relative frequencies are given for the track; *p*-values are derived by regressing the variable on a treatment indicator with standard errors clustered at school level, which are adjusted using the wild bootstrap approach; Financial scores are reported before standardization.

In terms of financial characteristics, students across all conditions scored, on average, 3.77 out of 8 on the pre-treatment test. Both in the *adaptive, elaborated feedback* and *adaptive, verification feedback* condition, students performed significantly worse in the behavior measure. We carefully assess the effect of the program via outcomes of the post-treatment test, as shown in panel D. The significant differences for the *uniform, elaborated feedback* and *adaptive, verification feedback* condition provide a first indication of a positive treatment effect. This finding should be interpreted with caution, however, given the baseline imbalances.

It should be noted that, while the final sample includes 1,177 students, 1,921 students initially took the pre-treatment test. We discuss the importance of attrition extensively in Appendix E. We find that attrition is selective and affects the validity of our study to some extent, i.e., in-sample schools included relatively more students from a favorable socioeconomic status (more native speaking students and fewer students with a mother without a secondary education degree) and non-complying students in the control condition scored lower on the baseline financial behavior measure, suggesting lower bound effects for this measure. Accordingly, to assess the extent to which the baseline imbalances and attrition

mirror subsequent estimations, we will examine the robustness of the following estimates using multiple tests.

5 Results

5.1 Main analysis

The first row in Table IV (*Treat*) presents the short-term effectiveness of the standard financial education program. The first three columns (corresponding to the three model specifications in which control variables and strata fixed effects are gradually added) present the estimates for students' financial proficiency. After controlling for the observed heterogeneity and stratification procedure (column 3), students' financial proficiency significantly improves by 0.47 standard deviations on average. This magnitude of improvement is similar to Hinojosa et al. (2010) who found an increase of 0.45 standard deviations for students in sixth to eighth grade and 0.39 standard deviations for students in ninth and tenth grade. In our most conservative estimation, i.e., when the inference is based on wild bootstrap to account for the few (treated) clusters in our data, the significance disappears.

Table IV: Effects on Short-Term Financial Proficiency

	Financial Proficiency			Financial Knowledge			Financial Behavior		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treat	0.651** (0.248)	0.476** (0.192)	0.469** (0.223)	0.697*** (0.220)	0.529*** (0.189)	0.485** (0.201)	0.357 (0.213)	0.223 (0.175)	0.178 (0.176)
<i>p</i> -value wild bootstrap	0.052*	0.030**	0.144	0.023**	0.014**	0.063*	0.177	0.288	0.458
adaptive	-0.562** (0.217)	-0.300* (0.166)	-0.180* (0.100)	-0.492** (0.194)	-0.290* (0.164)	-0.147* (0.0863)	-0.457** (0.179)	-0.219* (0.121)	-0.153* (0.0865)
<i>p</i> -value wild bootstrap	0.126	0.329	0.393	0.129	0.322	0.344	0.136	0.335	0.414
EF	-0.121 (0.207)	0.0602 (0.187)	-0.0852 (0.173)	-0.101 (0.180)	0.0480 (0.183)	-0.0694 (0.166)	-0.105 (0.179)	0.0758 (0.160)	-0.0132 (0.123)
<i>p</i> -value wild bootstrap	0.622	0.785	0.796	0.650	0.824	0.796	0.594	0.708	0.933
Strata fixed effects	No	No	Yes	No	No	Yes	No	No	Yes
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	1,177	1,177	1,177	1,177	1,177	1,177	1,177	1,177	1,177

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Table shows estimates from three model specifications for each outcome variable; Columns 1, 4, and 7 show the estimates from the basic specification, columns 2, 5, and 8 show the estimates from the specification including control variables, columns 3, 6, and 9 show the estimates from the specification including control variables and strata fixed effects; EF refers to elaborated feedback; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values; Controls: baseline financial value, grade in mathematics, track, gender, language spoken at home, and time between tests.

To gain insight into which financial component potentially drives the improvement in financial proficiency, we re-estimate the three models for financial knowledge and financial behavior separately. The results in Table IV indicate the short-term post-treatment performance to be primarily determined by an improvement in financial knowledge, i.e., students' financial knowledge increases by 0.49 standard deviations on average, whereas financial behavior is not significantly affected by the program. Accounting for the few clusters, estimating the standard errors by the wild bootstrap procedure still yields significant effect sizes for financial knowledge. Further, we find that also in the long term, this increase in financial knowledge is retained by students, as shown in Table C.I in Appendix C. Note,

however, that we must interpret this finding with caution due to the small and selective sample (see Figure E.I in Appendix E).

Via the interaction terms, we identify whether adaptive instruction (*adaptive*) and elaborated feedback (*EF*) enhance the program effectiveness as compared to a uniform learning path with verification feedback. Overall, we find that the practices have a negative, though limited, additional impact on the program effectiveness. Adapting the level of the learning path appears to deteriorate students' financial proficiency by 0.18 standard deviations. Providing the students with additional feedback reduces performance even further by 0.09 standard deviations, though these estimates are not statistically significant when inference is based on the wild bootstrap. This finding holds for the financial knowledge and behavior measure separately as well. Note that our main findings are robust against imbalance in covariates, attrition, and contamination of teacher characteristics, as examined using four tests in Appendix F.

5.2 Heterogeneity analysis

In spite of the absence of differentiation effects for the average student, research indicates the effects may be heterogeneous with respect to several student and classroom characteristics. Hence, we test for treatment heterogeneity next.

De Witte and van Klaveren (2014) showed that teachers typically teach at an intermediate level. Accordingly, we expect students with low- or high prior knowledge to do better when the learning path is adaptive, as it provides the opportunity to learn at the level students are functioning at. As demonstrated by Fyfe and Rittle-Johnson (2016), the effects of feedback can also be moderated by prior knowledge. In particular, they found that, while students with low prior knowledge learn more when receiving feedback, the reverse holds for students with high prior knowledge. We evaluate heterogeneity by students' baseline financial proficiency in Table C.II. The results show that adapting the level of the learning path does not significantly enhance the program effectiveness for all types of students (in the bottom, mid, or top tercile of the distribution), nor does providing an extended form of feedback.

Barrow et al. (2009) argued that the effectiveness of computer-assisted programs depends on class size and –heterogeneity. Differentiation is likely to be more effective in large and heterogeneous classes as greater variation in baseline proficiency makes it more challenging for teachers to design uniform material. Also, teachers need to spend their instructional time and time for feedback among more students in larger classes. Table C.III and Table C.IV present the estimates allowing the effects of the adaptive practices to differ by both class characteristics. While the standard program effectiveness appears lower, adaptive instruction and elaborated feedback significantly increase students' financial knowledge in larger classes by 0.03 and 0.05 standard deviations, respectively. Note, however, when inference is based on wild bootstrap, the results do not support this finding, neither in larger nor in more heterogeneous classes.

Narciss et al. (2014) find that gender mediates the effect of feedback. Despite elaborated feedback does not influence the performance of boys and girls differently in our study, we find significant heterogeneity for adaptive instruction, i.e., the performance of girls compared to boys is significantly reduced (at the ten percent level) when the learning path is adaptive, as presented in Table C.V.

5.3 Mechanism

It is widely argued that feedback both regulates and is regulated by the motivational state of students (e.g., Timmers et al., 2013; Timmers & Veldkamp, 2011). As suggested by Vollmeyer and Rheinberg (2005), feedback manipulation can affect students' motivation, which, in turn, is considered an important mediator for learning. Given the absence of learning gains by adaptive instruction and elaborated feedback in our study, we explore students' motivation during the lectures next.

Table V shows that adding the adaptive practices to the financial education program reduces the motivational state of students during the lectures. In particular, students' score on the motivation measure is 0.44 standard deviations lower when the learning path was adaptive, which deteriorates further by 0.65 standard deviations when the path included elaborated feedback. The effect sizes remain significant if we use the wild bootstrap approach.

Table V: Effects on Students' Motivational State

Dependent variable	Motivational State (1)
Adaptive	-0.444*** (0.0311)
<i>p</i> -value wild bootstrap	0.005***
EF	-0.646*** (0.0859)
<i>p</i> -value wild bootstrap	0.023**
Strata fixed effects	Yes
Controls	Yes
Observations	837

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; EF refers to elaborated feedback; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values; Controls: baseline financial proficiency, importance financial literacy, grade in mathematics, track, gender, and language spoken at home; Control condition excluded from analysis; Outcome measure standardized.

6 Discussion

The present paper examined the effectiveness of a computer-assisted financial education program in eighth- and ninth-grade students in Flemish schools (RQ1), and studied, in particular, the effects of adaptive instruction and elaborated feedback in the program (RQ2). Moreover, it explored whether class and student characteristics moderated the effects of the adaptive practices (RQ3).

First, as expected in research hypothesis H1, we found that the standard financial education program led to gains in students' financial proficiency. The improvement was mainly driven by financial knowledge, which increased by almost half of a standard deviation. We did not observe significant changes in students' financial behavior, which is in line with other financial education programs reporting only limited effect sizes for financial behavior (Kaiser & Menkhoff, 2019). The absence of an effect can potentially be attributed to the age of the target population, having too limited exposure to money.

Second, given the discrepancies in financial literacy found among youth (Lusardi et al., 2010), we expected that an adaptive version of the financial education program would enhance the learning outcomes of students (research hypothesis H2a). The results did not support this hypothesis, i.e., no main effects of the adaptive practices were documented. Yet, considering the effects of adaptive instruction and elaborated feedback were a priori ambiguous from the educational science literature (research hypothesis H2b), our results are in line with studies showing one-size-fits-all learning material to be as (or more) effective as adaptive learning material (van Klaveren et al., 2017).

Finally, the results related to the moderators of the adaptive practices partially supported research hypothesis H3. While the prior knowledge of students and the class size and –heterogeneity did not appear to influence the effects, a marginally significant heterogeneous effect for gender was reported, i.e., girls appeared to be negatively affected by adaptive instruction. Moreover, the results suggested that students' motivational state during the lectures influenced the effectiveness of the adaptive practices. Feedback is found to regulate students' motivation, which, in turn, affects their performance (Vollmeyer & Rheinberg, 2005). In our study, we found that offering adaptive instruction and elaborated feedback was linked to lower motivation levels. Consequently, it is possible that the adaptive practices did not affect their knowledge acquisition and thus, their final performance in the test. Shute (2008) argues that elaborated feedback should not be too long or complex for students to pay attention to it. As the learning path included already enhanced information effort (multiple information sheets and formative tests), it is possible that the supplementary provision of instruction and detailed feedback was perceived as 'excessive' by students, thereby reducing their motivation and rendering it useless. In turn, students' motivation may have affected feedback behavior, such as feedback-seeking and attention paid to feedback (Timmers & Veldkamp, 2011; Timmers et al., 2013). Note that, as girls are generally less interested in financial matters (e.g., Lührmann et al., 2015), this mechanism may have been amplified for them, resulting in a significantly negative effect of adaptive instruction on performance.

7 Conclusion

The findings of the present paper provide evidence that the gains of computer-assisted adaptive instruction and elaborated feedback, both in terms of performance and motivation, within education and financial education in particular, are likely much smaller than those claimed by researchers and policymakers. Considering the costs associated with adaptive technology tools (Shute & Zapata-Rivera,

2012), we contribute to the stream of literature that is critical towards adaptive practice programs showing uniform learning material to be as (or even more) effective.

7.1 Limitations

There are several reasons to be careful in extrapolating the effects of our study more broadly. First, there is a concern related to the external validity. Our sample of schools included relatively more students from a favorable socioeconomic status as compared to the average Flemish school. Consequently, this limitation is likely to overestimate the improvement in financial proficiency and potentially underestimated the impact of the adaptive practices, if we assume larger effects for students with lower socioeconomic status (as suggested for non-native students by Iterbeke et al., 2020).

As a second limitation, the financial proficiency test used in our study had low internal reliability. While this could point to the fact that the test items did not measure the same latent variable (financial proficiency), the low level of reliability may be attributed to the broad range of tested financial knowledge and behavior and the limited number of test items.

Third, as the study was conducted in a natural classroom setting, we had no control over the time spent on the learning path and, in particular, on the instruction and feedback offered in the learning path. While time spent on the task appears to be a moderator of feedback effectiveness (van der Kleij et al., 2012), we do not know in our study whether students effectively paid sufficient attention to it.

Fourth, to increase the scalability of our program, students completed the computer-assisted learning path in pairs of two. Even though students were ability-grouped and hence, all students received the learning path at an appropriate learning level, it is possible that the effects of the adaptive practices were different if students had worked individually. For instance, research suggests that the time spent on feedback depends on particular student characteristics, such as their attitudes and motivation (van der Kleij et al., 2012). Hence, if students worked individually rather than in pair, they might have spent more (or less) time on the adaptive instruction and elaborated feedback.

7.2 Implications for future research

We took care in designing our study to prevent potentially confounding influences, such as teachers, from affecting the estimated impacts. Despite we can thus be certain that the increased financial knowledge of students was not attributed to adaptive instruction or elaborated feedback, we were, however, unable to isolate the impact of other components of the software, such as the formative tests. Given the body of evidence showing the effectiveness of formative tests (e.g., Wang, 2008), part of the gain in students' financial knowledge may be attributed to these tests. The effectiveness of formative tests could be investigated in future field experiments.

An additional area for future research could be to examine the effectiveness of the adaptive practices in a setting where teachers play a more active role in guiding students through the computer-based learning environment. As research suggests that technology tools are most effective when an element of

face-to-face instruction is included (Haelermans & Ghysels, 2017; Tamim et al., 2011), it might be interesting, for instance, to explore the effect of teachers encouraging students to actively use the instruction and feedback offered in the computer-based learning environment.

Finally, the results on students' motivational state suggested that the provision of adaptive instruction and elaborated feedback (in combination with information sheets and formative tests) might have invoked an extraneous cognitive overload in students. More research is needed to determine the optimal amount of instruction and feedback in computer-based learning environments. The objective of the adaptive practices should be to adapt students' intrinsic load, taking care that the full cognitive load is restricted to their cognitive capacity (Hollender et al., 2010).

References

- Aleven, V., Mclaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on Adaptive Learning Technologies. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (pp. 522–560). Routledge.
- Attali, Y., & van der Kleij, F. (2017). Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers & Education, 110*, 154–169. <https://doi.org/10.1016/j.compedu.2017.03.012>.
- Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics, 123*, 1235–1264. <https://doi.org/10.1162/qjec.122.3.1235>.
- 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. <https://doi.org/10.1257/pol.1.1.52>.
- Blackwell, M., Iacus, S., King, G., & Porro, G. (2010). cem: Coarsened Exact Matching in Stata. *The Stata Journal, 9*(4), 524–546. <https://doi.org/10.1177/1536867X0900900402>.
- Bover, O., Hospido, L., & Villanueva, E. (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*.
- Bruhn, M., & McKenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics, 1*(4), 200–232. <https://doi.org/10.1257/app.1.4.200>.
- 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. <https://doi.org/10.1257/app.20150149>.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics, 90*(3), 414–427. <https://doi.org/10.1162/rest.90.3.414>.
- Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88–113. <https://doi.org/10.1016/j.edurev.2013.01.001>.
- Compen, B., De Witte, K., & Schelfhout, W. (2020). The impact of teacher engagement in an interactive webinar series on the effectiveness of financial literacy education. *British Journal of Educational Technology*. In Press. <https://doi.org/10.1111/bjet.13013>.
- 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., & Van Klaveren, C. (2014). How are teachers teaching? A nonparametric approach. *Education Economics, 22*(1), 3–23. <https://doi.org/10.1080/09645292.2011.560448>.
- Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education, 106*, 83–96. <https://doi.org/10.1016/j.compedu.2016.12.001>.
- 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*.
- Fyfe, E. R. (2016). Providing feedback on computer-based algebra homework in middle-school classrooms. *Computers in Human Behavior, 63*, 568–574. <https://doi.org/10.1016/j.chb.2016.05.082>.

- Fyfe, E. R., & Rittle-johnson, B. (2016). Feedback Both Helps and Hinders Learning: The Causal Role of Prior Knowledge. *Journal of Educational Psychology, 108*(1), 82–97. <https://doi.org/10.1037/edu0000053>.
- Golke, S., Dörfler, T., & Artelt, C. (2015). The impact of elaborated feedback on text comprehension within a computer-based assessment. *Learning and Instruction, 39*, 123–136. <https://doi.org/10.1016/j.learninstruc.2015.05.009>.
- Haelermans, C., & Ghysels, J. (2017). The effect of individualized digital practice at home on math skills - Evidence from a two-stage experiment on whether and why it works. *Computers & Education, 113*, 119–134. <https://doi.org/10.1016/j.compedu.2017.05.010>.
- Haelermans, C., Ghysels, J., & Prince, F. (2015). Increasing performance by differentiated teaching? Experimental evidence of the student benefits of digital differentiation. *British Journal of Educational Technology, 46*(6), 1161–1174. <https://doi.org/10.1111/bjet.12209>.
- Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses relating to achievement*. New York, NY: Routledge.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review Of Educational Research, 77*(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Heymans, P. J., Godaert, E., Elen, J., van Braak, J., & Goeman, K. (2018). MICTIVO2018. Monitor voor ICT-integratie in het Vlaamse onderwijs. Eindrapport van O&O-opdracht: Meting ICT-integratie in het Vlaamse onderwijs (MICTIVO). KU Leuven / Universiteit Gent.
- 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*.
- Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human-computer interaction. *Computers in Human Behavior, 26*(6), 1278–1288. <https://doi.org/10.1016/j.chb.2010.05.031>.
- Iacus, S., King, G., & Porro, G. (2008). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis, 20*(1), 1–24. <https://doi.org/10.1093/pan/mpr013>.
- Iterbeke, K., De Witte, K., Declercq, K., & Schelfhout, W. (2020). The Effect of Ability Matching and Differentiated Instruction in Financial Literacy Education. Evidence from Two Randomised Control Trials. *Economics of Education Review, In Press*. <https://doi.org/10.1016/j.econedurev.2019.101949>.
- Kaiser, T., & Menkhoff, L. (2019). Financial education in schools: A meta-analysis of experimental studies. *Economics of Education Review, In Press*. <https://doi.org/10.1016/j.econedurev.2019.101930>.
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research, 86*(1), 42–78. <https://doi.org/10.3102/0034654315581420>.
- Kulik, C. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of Mastery Learning Programs: A Meta-Analysis. *Review Of Educational Research, 60*(2), 265–299. <https://doi.org/10.3102/00346543060002265>.
- Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment effects. *Review of Economic Studies, 76*, 1071–1102. <https://doi.org/10.1111/j.1467-937X.2009.00536.x>.
- Lusardi, A., & Mitchell, O. S. (2011). Financial literacy and planning: Implications for retirement wellbeing. *NBER Working Paper Series No 17078*.
- Lusardi, A., Mitchell, O. S., & Curto, V. (2010). Financial Literacy among the Young. *The Journal of Consumer Affairs, 44*(2), 358–380. <https://doi.org/10.1111/j.1745-6606.2010.01173.x>.

- Lührmann, M., Serra-Garcia, M., & Winter, J. (2015). Teaching teenagers in finance: Does it work? *Journal of Banking and Finance*, *54*, 160–164. <https://doi.org/10.1016/j.jbankfin.2014.11.009>.
- 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. <https://doi.org/10.1257/pol.20170012>.
- MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal*, *21*, 114–135. <https://doi.org/10.1111/ectj.12107>.
- Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education*, *95*, 85–98. <https://doi.org/10.1016/j.compedu.2015.12.002>.
- Maldonado, J., De Witte, K. and Declercq, K. (2019). The effects of parental involvement in homework. Two randomised controlled trials in financial education. *Department of Economics KU Leuven Discussion Paper Series 19.14*. pp. 68.
- Mory, E. H. (2004). Feedback research revisited. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745–784). Mahwah: Erlbaum Associates.
- Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. *American Economic Review*, *109*(4), 1426–1460. <https://doi.org/10.1257/aer.20171112>.
- Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, G., & Melis, E. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, *71*, 56–76. <https://doi.org/10.1016/j.compedu.2013.09.011>.
- 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.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & Wilbert J. McKeachie. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*, 801–813. <https://doi.org/10.1177/0013164493053003024>.
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, *78*(1), 153–189. <https://doi.org/10.3102/0034654307313795>.
- Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive Educational Systems. In P. J. Durlach & A. M. Lesgold (Eds.), *Adaptive Technologies for Training and Education* (pp. 7–27). New York: Cambridge University Press.
- Sottolare, R., & Goldberg, B. (2012). Designing Adaptive Computer-Based Tutoring Systems to Accelerate Learning and Facilitate Retention. *Cognitive Technology*, *17*(1), 19–34.
- Subban, P. (2006). Differentiated instruction: A research basis. *International Education Journal*, *7*(7), 935–947. <https://doi.org/10.1111/j.1365-2648.2006.04074.x>.
- Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study. *Review Of Educational Research*, *81*(1), 4–28. <https://doi.org/10.3102/0034654310393361>.
- Timmers, C. F., Braber-van den Broek, J., & van den Berg, S. M. (2013). Motivational beliefs, student effort, and feedback behaviour in computer-based formative assessment. *Computers & Education*, *60*, 25–31. <https://doi.org/10.1016/j.compedu.2012.07.007>.
- Timmers, C., & Veldkamp, B. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers & Education*, *56*, 923–930. <https://doi.org/10.1016/j.compedu.2010.11.007>.

- Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., Conover, L. A., & Reynolds, T. (2003). Differentiating Instruction in Response to Student Readiness, Interest, and Learning Profile in Academically Diverse Classrooms: A Review of Literature. *Journal for the Education of the Gifted*, 27(2–3), 119–145. <https://doi.org/10.1177/016235320302700203>.
- Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, 27, 118–130. <http://dx.doi.org/10.1016/j.chb.2010.07.038>.
- Van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58, 263–272. <https://doi.org/10.1016/j.compedu.2011.07.020>.
- Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis. *Review of Educational Research*, 85(4), 475–511. <https://doi.org/10.3102/0034654314564881>.
- van Klaveren, C., Vonk, S., & Cornelisz, I. (2017). The effect of adaptive versus static practicing on student learning - evidence from a randomized field experiment. *Economics of Education Review*, 58, 175–187. <https://doi.org/10.1016/j.econedurev.2017.04.003>.
- Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction*, 15, 589–602. <https://doi.org/10.1016/j.learninstruc.2005.08.001>.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wang, T. (2008). Web-based quiz-game-like formative assessment: Development and evaluation. *Computers & Education*, 51, 1247–1263. <https://doi.org/10.1016/j.compedu.2007.11.011>.
- Wang, Z., Gong, S., Xu, S., & Hu, X. (2019). Elaborated feedback and learning: Examining cognitive and motivational influences. *Computers & Education*, 136, 130–140. <https://doi.org/10.1016/j.compedu.2019.04.003>.

Appendix

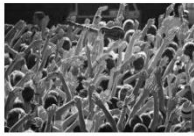
Appendix A: Context and curriculum

This paper focuses on Flanders, the northern region of Belgium. The Flemish education system comprises three educational networks, i.e., publicly funded education managed by the Community authorities, grant-aided education managed by other levels of the government, and grant-aided private education, which is the largest network. Despite receiving comparable grants, private education includes, on average, students with a higher socioeconomic status. Flemish secondary education is organized in a tracking system. In particular, students in eighth grade are tracked in an academic or pre-vocational education track, whereas students in ninth grade in an academic, technical, arts, or vocational education track. Within this structure of ability tracking, our intervention of adaptive instruction and feedback provides a more fine-grained way of differentiation and deepens the prevailing between-classroom differentiation of ability tracking.

The PISA financial literacy assessment indicates that the Flemish student performs above average. Yet, large discrepancies in students' performances can be found within the region. In particular, large gaps at the bottom end of the performance distribution are prevailing, both in absolute terms and relative to the gaps at the top end (OECD, 2017). It should be noted that financial competencies were not included in the Flemish education curriculum during our study. However, as they became part of the education program from September 2019 onwards, teachers were indirectly incentivized to participate in our study.

Flemish schools are well equipped with ICT as nearly all secondary schools have internet access, and schools have, on average, 250 computers available for their students (Heymans, Godaert, Elen, van Braak, & Goeman, 2018). However, the actual use of ICT is limited and estimated to be once or several times a month. Computer-assisted instruction is even less frequently used (only several times a year) for classroom differentiation (Heymans et al., 2018).

Appendix B: Figures

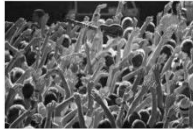


Sarah saves € 28,00 of her pocket money every month.
 This summer, she plans to go to a music festival with her friends.
 A combi ticket for four days festival and access to the camping site costs € 140,00.
 Calculate how many months it will take Sarah to save in order to buy the ticket.
 If Sarah wants to go to the festival, she needs to save for months.



If you find this exercise difficult, you may click on the tip.

(a) Exercise in the Basic-Level Learning Path

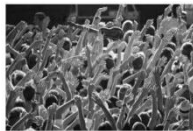


Sarah saves € 15,00 of her pocket money every month.
 This summer, she plans to go to a music festival with her friends.
 A combi ticket for four days festival and access to the camping site costs € 140,00.
 Calculate how many months it will take Sarah to save in order to buy the ticket.
 Assume a month counts 30 days. Round up the number of days.
 If Sarah wants to go to the festival, she needs to save for months and days.



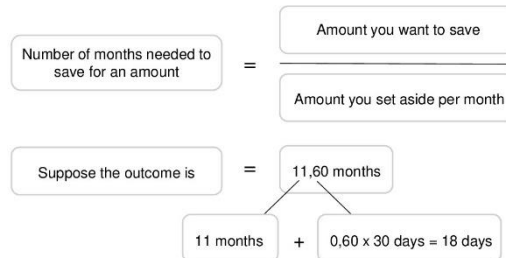
Use the tip if you can't find the answer.

(b) Exercise in the Intermediate-Level Learning Path



Sarah saves € 15,00 of her pocket money every month.
 This summer, she plans to go to a music festival with her friends.
 A combi ticket for four days festival and access to the camping site costs € 140,00.
 Calculate how many months it will take Sarah to save in order to buy the ticket.
 Assume a month counts 30 days. Round up the number of days.
 If Sarah wants to go to the festival, she needs to save for months and days.

(c) Exercise in the Advanced-Level Learning Path



(d) Tip in the Intermediate-Level Learning Path

Figure B.I: Differences in Instruction in the Adaptive Learning Path

Question 1 of 3

What is price stability in the euro area?

- Prices do not increase
- Inflation
- Maintain inflation rates below, but close to, 2%
- A decrease of the general price level

Submit

Next

Feedback

This answer is incorrect.

(a) *Verification Feedback*

Question 1 of 3

What is price stability in the euro area?

- Prices do not increase
- Inflation
- Maintain inflation rates below, but close to, 2%
- A decrease of the general price level

Submit

Next

Feedback

A decrease of the general price level is deflation.

Price stability is defined as maintaining inflation rates below, but close to, 2%.

This answer is incorrect.

(b) *Elaborated Feedback*

Figure B.II: *Verification versus Elaborated Feedback in the Learning Path*

Appendix C: Tables

Table C.I: Effects on Long-Term Financial Proficiency

	Financial Proficiency		Financial Knowledge		Financial Behavior	
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	0.0814	0.399**	0.207	0.497***	-0.120	0.193
	(0.244)	(0.184)	(0.204)	(0.161)	(0.232)	(0.209)
<i>p</i> -value wild bootstrap	0.784	0.127	0.312	0.068*	0.726	0.464
adaptive	-0.0620	0.0360	-0.119	-0.0922	0.0390	0.0894
	(0.178)	(0.198)	(0.153)	(0.165)	(0.154)	(0.173)
<i>p</i> -value wild bootstrap	0.722	0.787	0.436	0.647	0.790	0.645
EF	0.0436	0.134	0.110	0.229**	-0.0635	0.00859
	(0.209)	(0.110)	(0.156)	(0.0907)	(0.218)	(0.151)
<i>p</i> -value wild bootstrap	0.890	0.464	0.605	0.203	0.834	0.982
Strata fixed effects	No	No	No	No	No	No
Controls	No	Yes	No	Yes	No	Yes
Observations	544	544	544	544	544	544

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Table shows estimates from two model specifications for each outcome variable; Columns 1, 3, and 5 show the estimates from the basic specification, columns 2, 4, and 6 show the estimates from the specification including control variables; EF refers to elaborated feedback; Strata fixed effects are excluded due to multicollinearity issues; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values; Controls: baseline financial value, grade in mathematics, track, gender, language spoken at home, and time between tests; We use the performance in the first post-test to measure long-term performance of students in the control condition.

Table C.II: Heterogeneity by Baseline Financial Proficiency Terciles

Dependent variable	Financial Proficiency	Financial Knowledge	Financial Behavior
	(1)	(2)	(3)
Treat * bottom tercile	-0.265	-0.225	-0.224
	(0.172)	(0.199)	(0.201)
<i>p</i> -value wild bootstrap	0.205	0.347	0.288
Treat * top tercile	-0.0578	-0.183	0.134
	(0.239)	(0.210)	(0.252)
<i>p</i> -value wild bootstrap	0.827	0.424	0.605
adaptive * bottom tercile	0.0312	-0.0178	0.0869
	(0.143)	(0.139)	(0.129)
<i>p</i> -value wild bootstrap	0.834	0.899	0.528
adaptive * top tercile	0.127	0.213	-0.0355
	(0.199)	(0.136)	(0.232)
<i>p</i> -value wild bootstrap	0.542	0.222	0.901
EF * bottom tercile	0.164	0.0610	0.246
	(0.158)	(0.175)	(0.152)
<i>p</i> -value wild bootstrap	0.357	0.772	0.120
EF * top tercile	-0.0435	0.0304	-0.129
	(0.180)	(0.171)	(0.178)
<i>p</i> -value wild bootstrap	0.828	0.871	0.498
Strata fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	1,177	1,177	1,177

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; EF refers to elaborated feedback; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values; Controls: grade in mathematics, track, gender, language spoken at home, and time between tests; Reference: middle tercile.

Table C.III: Heterogeneity by Class Size

Dependent variable	Financial Proficiency	Financial Knowledge	Financial Behavior
	(1)	(2)	(3)
Treat * Class size	-0.0313 (0.0248)	-0.0669** (0.0261)	0.0320 (0.0416)
<i>p</i> -value wild bootstrap	0.190	0.044**	0.630
adaptive * Class size	0.0215* (0.0113)	0.0344* (0.0189)	-0.00621 (0.0264)
<i>p</i> -value wild bootstrap	0.165	0.216	0.799
EF * Class size	0.0187 (0.0255)	0.0522** (0.0203)	-0.0310 (0.0318)
<i>p</i> -value wild bootstrap	0.496	0.109	0.600
Strata fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	1,177	1,177	1,177

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; EF refers to elaborated feedback; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values; Controls: baseline financial value, grade in mathematics, track, gender, language spoken at home, and time between tests.

Table C.IV: Heterogeneity by Class Baseline Financial Proficiency Standard Deviation

Dependent variable	Financial Proficiency	Financial Knowledge	Financial Behavior
	(1)	(2)	(3)
Treat * Class baseline SD	0.182 (0.555)	0.312 (0.588)	-0.236 (0.436)
<i>p</i> -value wild bootstrap	0.780	0.648	0.614
adaptive * Class baseline SD	0.00190 (0.311)	-0.272 (0.294)	0.379 (0.291)
<i>p</i> -value wild bootstrap	0.995	0.347	0.333
EF * Class baseline SD	-0.225 (0.406)	-0.0553 (0.432)	-0.389 (0.350)
<i>p</i> -value wild bootstrap	0.645	0.918	0.371
Strata fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	1,177	1,177	1,177

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; EF refers to elaborated feedback; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values; Controls: baseline financial value, grade in mathematics, track, gender, language spoken at home, and time between tests.

Table C.V: Heterogeneity by Gender

Dependent variable	Financial Proficiency (1)	Financial Knowledge (2)	Financial Behavior (3)
Treat * Female	0.165	0.341	-0.187
	(0.202)	(0.228)	(0.203)
<i>p</i> -value wild bootstrap	0.439	0.185	0.430
adaptive * Female	-0.271**	-0.299*	-0.0989
	(0.120)	(0.161)	(0.145)
<i>p</i> -value wild bootstrap	0.091*	0.238	0.522
EF * Female	-0.0240	-0.162	0.195
	(0.128)	(0.125)	(0.132)
<i>p</i> -value wild bootstrap	0.861	0.275	0.184
Strata fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	1,177	1,177	1,177

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; EF refers to elaborated feedback; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values; Controls: baseline financial value, grade in mathematics, track, language spoken at home, and time between tests.

Appendix D: Test instruments

D.1 First Post-Treatment Financial Proficiency Test

1. *You receive € 60,00 pocket money every month. You pay € 10,00 every two weeks at your tennis club and € 20,00 monthly for your mobile phone service. One month consists of four weeks. How much money do you have left every month?*
 - € 38,00
 - € 28,00
 - € 40,00
 - € 8,00
 - I don't know
2. *John receives € 50,00 from his grandmother for his birthday. He deposits the amount on his savings account. The interest on his savings account amounts to 2% per year. The inflation amounts to 1% per year. After one year, John can buy:*
 - More than today
 - Less than today
 - As much as today
 - I don't know
3. *You open up a savings account and deposit € 200,00. The interest on the savings account amounts to 2% per year. How much money will be on your savings account after five years, if you do not withdraw or deposit additional amounts:*
 - Less than € 220,00
 - Exactly € 220,00
 - More than € 220,00
 - I don't know
4. *You receive € 100,00 from your parents for your birthday. Which of the following options will give you the highest return?*
 - You deposit the money on your savings account
 - You deposit the money on your current account
 - You keep the money in your wallet
 - I don't know
5. *You own an amount of money which you don't need in the near future. Which option makes the most sense?*
 - You deposit the money on a savings account.
 - You put your money in your piggy bank.
 - You keep the money in your wallet.
 - You deposit the money on a current account.
 - I don't know
6. *Which of the following forms of saving and investment will give you the lowest risk of losing everything you put into it, but also the lowest return?*
 - A bond
 - A savings account at a bank
 - A share of an enterprise
 - An investment trust
 - I don't know
7. *Bart's bike was stolen last week. He has to buy a new one, but he has no savings and his parents cannot lend him any money. He wants to start saving. However, he does not know how to proceed. What would you do if you were him?*

- You put all your pocket money in a piggy bank.
 - You deposit the money you have left at the end of the month on a savings account.
 - You set up a budget with your incomes and expenses, calculate how much you can save every month, and put the amount on a savings account.
 - Forget the bike! With your high expenses for games, you will never be able to save.
 - I don't know
8. *You are contacted by a foreign investment trust with an interesting offer. They offer you to invest your money in their trust with a guaranteed return of 25%. This is much higher than the 4% you now get at your Belgian investment trust. How would you respond to this offer?*
- The offer sounds too good to be true. This can't be realistic and I would not take the offer.
 - I would put half of my savings into the trust. If I spread my money this way, I lower my risks.
 - I would search for additional information on their website. If the numbers appear correct, I would take the offer.
 - I would definitely take the offer.
 - I don't know

Appendix E: Attrition

E.1 Internal Validity

The differences observed in Table III may result from attrition, with two main sources in the present study. First, teachers registered with 180 classes, however, 44 (24 percent) classes dropped out of the study prior to taking the pre-treatment test. Although this is only a weak form of attrition, these classes matter for the design as the stratified randomization included these classes. Second, while 1,921 students completed the pre-treatment test, 744 (39 percent) students did not take the first post-test. We examine the attrition rates more in-depth and find that 73 percent of attrition can be explained by teachers not following the prescribed instructions, i.e., the post-test was not administered in 42 classes.

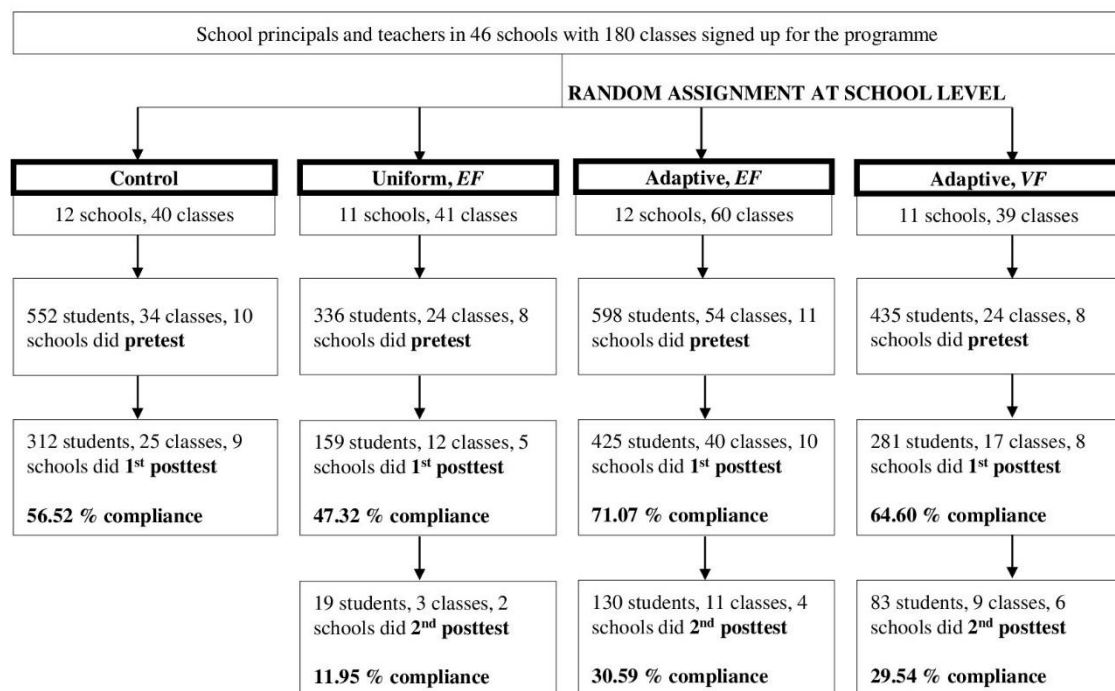


Figure E.I: Allocation of Students

Figure E.I indicates the attrition rate for the first post-test to be distinctly higher for the *uniform, elaborated feedback* condition. We can test for selective attrition between the pre- and first post-test across the different conditions using the baseline information that we collected for all students. By regressing an attrition indicator (i.e., a dummy indicating missing values for the post-treatment outcome) on treatment indicators, controls, and strata fixed effects, the presence of selective attrition cannot be ruled out, as shown in Table E.I. We find that students following the adaptive learning path during the lectures were significantly more likely to complete the first post-test. Note that, as the protocols in all experimental conditions were very similar, we believe this pattern of attrition was not caused by design. Further, Table E.II compares complying (i.e., students who took both tests) and non-complying students (i.e., students who only took the pre-test) within each condition. Apart from some small differences in the experimental conditions, non-complying students in the control condition appeared to score

significantly lower on the baseline financial behavior measure than complying students, suggesting lower bound effects for this measure.

Table E.I: Selective Attrition

	Attrition		
	(1)	(2)	(3)
Treat	0.157	0.185	0.123
	(0.223)	(0.210)	(0.158)
<i>p</i> -value wild bootstrap	0.554	0.459	0.560
adaptive	-0.237	-0.262	-0.288***
	(0.182)	(0.169)	(0.0913)
<i>p</i> -value wild bootstrap	0.291	0.192	0.032**
adaptive * EF	-0.0647	-0.0980	0.0292
	(0.106)	(0.111)	(0.148)
<i>p</i> -value wild bootstrap	0.563	0.438	0.850
Strata fixed effects	No	No	Yes
Controls	No	Yes	Yes
Observations	1,921	1,921	1,921

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; Table shows estimates from three model specifications; Column 1 shows the estimates from the basic specification, column 2 shows the estimates from the specification including control variables, column 3 shows the estimates from the specification including control variables and strata fixed effects; EF refers to elaborated feedback; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values; Controls: baseline financial value, grade in mathematics, track, gender, language spoken at home.

Table E.II: Difference between Complying and Non-complying Students

Variables	Control	<i>p</i> - value	Uniform, <i>EF</i>	<i>p</i> - value	Adaptive, <i>EF</i>	<i>p</i> - value	Adaptive, <i>VF</i>	<i>p</i> - value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Percentage of Attrition</i>	43.48		52.68		28.93		35.40	
<i>Number of Attrited Students</i>	240		177		173		154	
Panel A. School characteristics								
Private	-0.12 (0.16)	0.507	-0.36 (0.28)	0.414	-0.08 (0.08)	0.258	0.08 (0.08)	0.367
Class size	0.55 (1.43)	0.730	-3.27 (1.72)	0.063	-1.07 (2.47)	0.776	-0.84 (0.96)	0.516
Fraction of 8 th grade students	0.12 (0.12)	0.328	0.03 (0.27)	0.906	-0.06 (0.20)	0.793	0.02 (0.11)	0.852
Panel B. Background characteristics								
Track								
Academic	-3.75%		1.26%		-6.71%		0.97%	
Technical	-1.91%	0.697	-4.92%	0.789	-7.98%	0.239	-0.20%	0.523
Vocational	5.66%		3.65%		14.99%		-0.77%	
Gender (female)	-0.09 (0.12)	0.654	0.08 (0.12)	0.555	-0.13 (0.04)	0.018	0.02 (0.05)	0.859
Age (years)	0.01 (0.14)	0.979	0.13 (0.28)	0.766	0.15 (0.24)	0.597	-0.07 (0.13)	0.633
Language (Dutch)	-0.10 (0.06)	0.182	-0.14 (0.12)	0.547	-0.13 (0.05)	0.040	0.02 (0.04)	0.727
Grade in mathematics (5-points)	-0.03 (0.16)	0.866	-0.19 (0.16)	0.344	0.02 (0.16)	0.903	0.15 (0.19)	0.453
Importance financial literacy (5-points)	-0.10 (0.11)	0.475	-0.03 (0.09)	0.797	-0.16 (0.10)	0.251	-0.08 (0.06)	0.227
Panel C. Pre financial scores								
Financial proficiency (9-points)	-0.19 (0.26)	0.490	-0.53 (0.39)	0.258	-0.11 (0.28)	0.680	0.23 (0.26)	0.375
Financial knowledge (5-points)	0.09 (0.20)	0.766	-0.26 (0.17)	0.148	0.03 (0.17)	0.894	-0.03 (0.17)	0.883
Financial behavior (3-points)	-0.28 (0.09)	0.039	-0.27 (0.24)	0.391	-0.14 (0.13)	0.318	0.26 (0.12)	0.164

Note: Differences in means between complying and non-complying students are obtained from regressing the characteristic on an attrition indicator with standard errors clustered at the school level (in parentheses); EF refers to elaborated feedback, VF refers to verification feedback; *p*-values are adjusted using the wild bootstrap approach; Relative differences are given for the track.

E.2 External Validity

To assess the representativeness of the schools in our sample, Table E.III compares school characteristics of in-sample and out-of-sample schools. In particular, four socio-economic indicators are examined using administrative data on all Flemish secondary schools, i.e., the percentage of children with a mother without a secondary education degree, the percentage of non-native children, the percentage of children receiving an allowance, and the percentage of children living in a neighborhood with high retention rates (defined as students whose study falls two years behind schedule at the age of fifteen).⁷ Overall, our sample represents secondary schools in Flanders fairly well, except for two socio-economic indicators, i.e., the in-sample schools include a lower share of non-native students and students having a low-educated mother as compared to the out-of-sample schools.

⁷ AGODI, Cijfermateriaal - Leerlingenkenmerken (2017-2018), available at <http://www.agodi.be/cijfermateriaal-leerlingenkenmerken>.

Table E.III: External Validity

Characteristic	In-sample Schools	Out-of-sample Schools	<i>p</i>- value
	(1)	(2)	(3)
% low educated mothers	18.95	25.10	0.027
% on allowance	24.71	28.71	0.100
% non-native	10.16	16.93	0.019
% neighborhood high retention	19.86	24.82	0.177

Note: *t*-tests are computed to derive the mean values and *p*-values; In-sample schools are defined as schools where students took the pre- and first post-treatment test; Four missing participating schools in the administrative data.

Appendix F: Robustness Tests

Using the following five tests, we assess whether our main results are robust against confounding variables, imbalance in covariates, attrition, and the contamination of teacher characteristics.

F.1 Confounding Variables

To formally test the absence of confounding variables, we compare the program effectiveness for average-performing students following the intermediate-level learning path in either the *uniform, elaborated feedback* condition or the *adaptive, elaborated feedback* condition. Given both experimental conditions *by definition* are equal for those students (i.e., students are offered an intermediate-level learning path with elaborated feedback), we expect them to perform equally well in the post-test. Students were asked which path they had followed in the post-test. Note that, as not all students completed the full test up to the questions related to the evaluation of the learning material, this analysis is restricted to a subgroup of students. To evaluate whether high-performing students did better following the intermediate- or advanced-level path, whether average-performing students performed similarly in the *adaptive, elaborated feedback* and *uniform, elaborated feedback* condition, and whether the performance of low-performing students differed in the basic or intermediate-level path, we generate counterfactual students in the *uniform, elaborated feedback* condition for students following a particular level in the *adaptive, elaborated feedback* condition. To do so, students are matched based on several characteristics and the model is re-estimated with the corresponding samples.

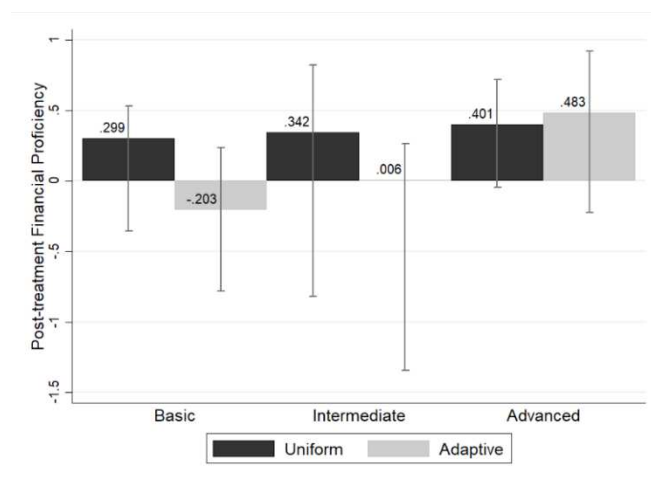


Figure F.I: Post-treatment Financial Proficiency by Level of Learning Path

Note: This figure shows the post-treatment financial proficiency for low-, average-, and high-performing students following a uniform or adaptive learning path (basic, intermediate, or advanced). A Coarsened Exact Matching approach by Iacus et al. (2008) and Blackwell et al. (2009) is used to match students in the *uniform, elaborated feedback* and *adaptive, elaborated feedback* condition. Students are matched based on baseline financial proficiency score, importance financial literacy, gender, language spoken at home, track, and grade in mathematics. Estimates are derived from a regression of the post-treatment financial proficiency score on a treatment indicator with bootstrapped standard errors clustered at school level using matched samples.

Figure F.I illustrates that for all levels of the learning path, the post-treatment proficiency of low-, average-, and high-performing students is not significantly different in both conditions. Accordingly, this finding implies the non-existence of confounding variables.

F.2 Baseline Imbalance

As we have repeated measurement for the main outcome variables, an alternative controlling for the baseline differences across experimental conditions and control condition is a classic Difference-in-Differences (DiD) model. The regression model is defined as follows:

$$y_{ist} = \alpha + \beta_0 post_{it} * Treat_{is} + \beta_1 post_{it} * adaptive_{is} + \beta_2 post_{it} * EF_{is} + \beta_3 Treat_{is} + \beta_4 adaptive_{is} + \beta_5 EF_{is} + \beta_6 post_{it} + \epsilon_{is} \quad (1)$$

where y_{ist} refers to the pre- and post-treatment values of the outcome measure. The variable $post_{it}$ takes value one for post-treatment observations (zero otherwise). The interaction terms $post_{it} * Treat_{is}$, $post_{it} * adaptive_{is}$, and $post_{it} * EF_{is}$ account for the treatment effects.

Table F.I: Difference-in-Differences

Dependent variable	Financial Proficiency (1)	Financial Knowledge (2)	Financial Behavior (3)
Treat	0.690*** (0.210)	0.590*** (0.193)	0.539*** (0.184)
<i>p</i> -value wild bootstrap	0.052*	0.039**	0.070*
adaptive	-0.297 (0.192)	-0.290* (0.169)	-0.231 (0.174)
<i>p</i> -value wild bootstrap	0.423	0.268	0.619
EF	-0.140 (0.132)	-0.0745 (0.148)	-0.167* (0.0893)
<i>p</i> -value wild bootstrap	0.388	0.699	0.106
Strata fixed effects	No	No	No
Controls	No	No	No
Observations	2,354	2,354	2,354

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; EF refers to elaborated feedback; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values.

The results in Table F.I confirm that the baseline imbalances are relatively unimportant, i.e., the estimates for the overall financial proficiency and financial knowledge measures from the DiD model are similar to the Intent-to-Treat ones. They nearly coincide with those from the first specification without strata fixed effects and controls. For the financial behavior measure, on the other hand, the average treatment effect is larger and significant in the DiD specification. Both the differentiation practices remain, on average, ineffective for teaching financial education.

F.3 Attrition

Inverse probability weighting (IPW) is often used to minimize bias due to differential attrition. Given we observe significant differences in attrition patterns across conditions, it is important to examine to what extent our results are robust against attrition. Using IPW, observations in control and experimental conditions are reweighted to remain comparable on important observed characteristics. It should be noted, however, that IPW relies on the assumption that we have enough information about students in the different conditions, so that, controlling for the observed characteristics, it is guaranteed that students are comparable on unobserved characteristics as well. Given this rather strong assumption, we further examine the importance of attrition using an alternative approach, i.e., Lee (2009) bounds, for which the assumption of selection on observables can be relaxed. To compute the trimming fractions and corresponding bounds, we use the residuals from a regression of the outcome on the baseline value of the outcome and control variables with standard errors clustered at school level. Moreover, we generate different sets of upper and lower bounds using shares of the actual trimming fractions (from 10 to 100 percent). This way, we are able to identify the exact level at which the estimates are robust.

Table F.II: Inverse Probability Weighting

Dependent variable	Financial Proficiency (1)	Financial Knowledge (2)	Financial Behavior (3)
Treat	0.479 (0.309)	0.607** (0.261)	0.130 (0.255)
<i>p</i> -value wild bootstrap	0.249	0.063*	0.693
adaptive	-0.154 (0.115)	-0.167 (0.0985)	-0.0702 (0.0981)
<i>p</i> -value wild bootstrap	0.300	0.188	0.560
EF	-0.108 (0.224)	-0.151 (0.198)	-0.00881 (0.172)
<i>p</i> -value wild bootstrap	0.769	0.636	0.975
Strata fixed effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	1,177	1,177	1,177

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; EF refers to elaborated feedback; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values; Controls: baseline financial value, grade in mathematics, track, gender, language spoken at home, and time between tests; Estimates are weighted by the inverse of the predicted probability of having non-missing data at endline; Probability is predicted by a multinomial logit model with baseline financial proficiency value, track, gender, age, language spoken at home, grade in mathematics, type of education, and the importance of saving as independent variables.

First, Table F.II shows that, when we account for inverse probability weights, the estimated Intent-to-Treat effects remain almost unchanged. Second, Figure F.II illustrates that for all trimming portions, the confidence intervals of the estimate and the bounds for the average treatment effect overlap, and that both the upper and lower bounds are positive. Accordingly, the estimate appears robust. The estimates for treatment heterogeneity by adapting the level of the learning path and providing elaborated feedback appear relatively robust as well, though the bounds are wider for the former.

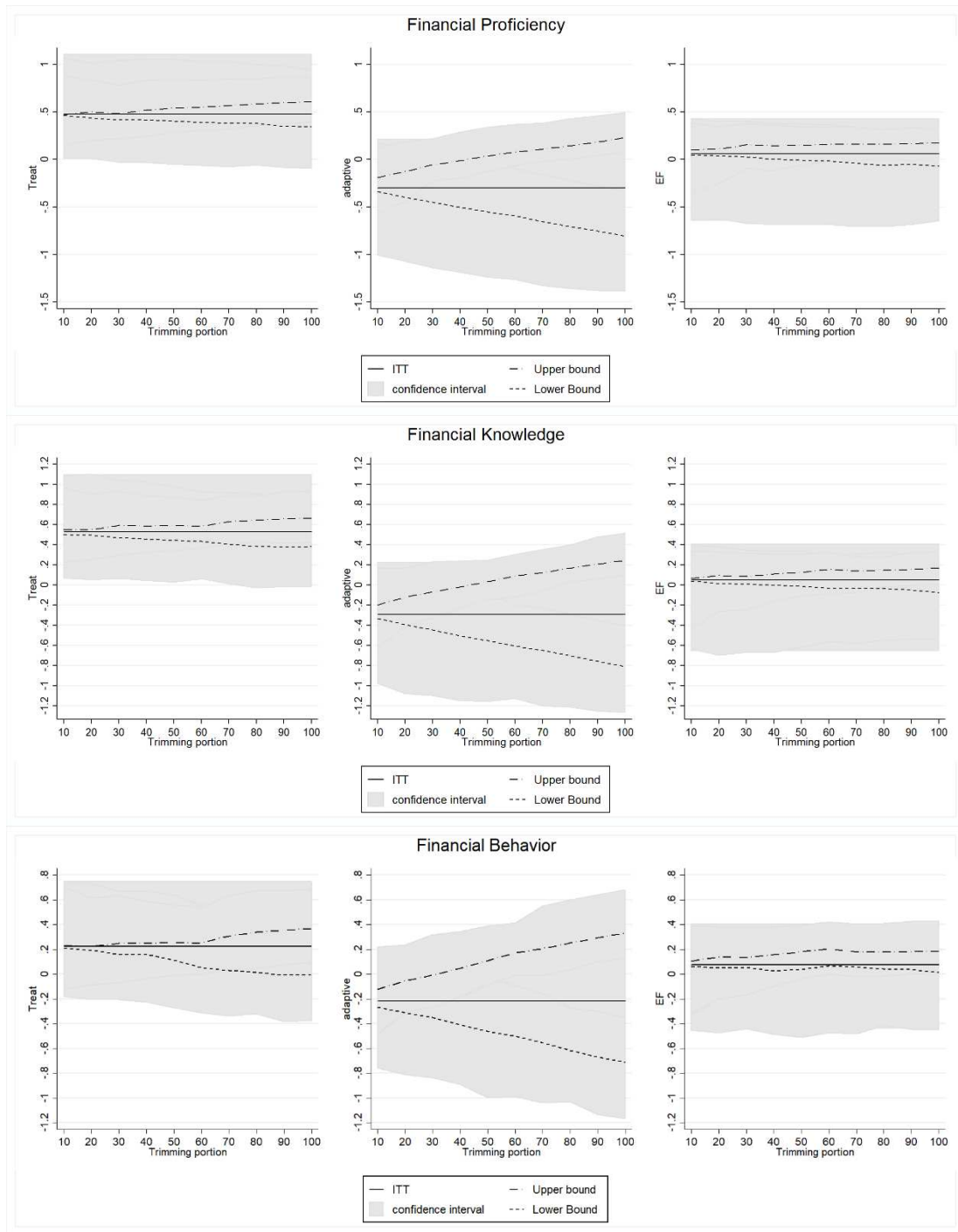


Figure F.II: Lee Bounds

Note: To compute the bounds, we first estimate the residuals from a regression of the outcome on the baseline value of the outcome and control variables with standard errors clustered at school level. The residuals are used to manually trim the conditions with the lowest attrition rates to equalize response rates across all conditions (see Figure E.I). We define different sets of bounds using shares of the actual trimming fractions (from 10 to 100 percent). The graphs illustrate for all three outcome measures the ITT estimate in solid line and the bounds in dashed lines with confidence intervals plotted against the different shares of the actual trimming fractions. Strata fixed effects are excluded and the inference is based on wild bootstrap.

F.4 Teacher Intervention

Despite the program design made a teacher intervention uncalled for, it is important to evaluate whether the effectiveness of the program is not contaminated by particular teacher characteristics. For

instance, in the event teachers helped students with the learning path, the effect of differentiation may be biased. Hence, we test this assumption using a regression model where potential teacher intervention is estimated by interacting the treatment indicators with teacher quality, which is measured via teachers' baseline financial proficiency, teaching experience, experience in teaching financial education, and self-efficacy for teaching financial education.

The results in Table F.III reveal that the treatment effects are orthogonal to teacher characteristics as the coefficients on the interactions are not jointly significant ($p = 0.502$). Hence, the observed treatment effect is not contaminated by particular teacher characteristics.

Table F.III: Heterogeneity by Teacher Characteristics

Dependent variable	Financial Proficiency (1)
Treat * baseline financial score	-0.164 (0.123)
<i>p</i> -value wild bootstrap	0.407
adaptive * baseline financial score	0.196** (0.0768)
<i>p</i> -value wild bootstrap	0.628
EF * baseline financial score	0.114 (0.121)
<i>p</i> -value wild bootstrap	0.495
Treat * years of experience	-0.0269** (0.0128)
<i>p</i> -value wild bootstrap	0.091*
adaptive * years of experience	0.0278** (0.0104)
<i>p</i> -value wild bootstrap	0.059*
EF * years of experience	0.0151 (0.0111)
<i>p</i> -value wild bootstrap	0.292
Treat * experience teaching FE	-0.131 (0.497)
<i>p</i> -value wild bootstrap	0.831
adaptive * experience teaching FE	-0.183 (0.391)
<i>p</i> -value wild bootstrap	0.675
EF * experience teaching FE	0.272 (0.350)
<i>p</i> -value wild bootstrap	0.569
Treat * self-efficacy teaching FE	0.115 (0.159)
<i>p</i> -value wild bootstrap	0.529
adaptive * self-efficacy teaching FE	0.137 (0.114)
<i>p</i> -value wild bootstrap	0.490
EF * self-efficacy teaching FE	-0.186* (0.0958)
<i>p</i> -value wild bootstrap	0.266
<i>p</i>-value of joint significance of interaction terms	0.502
Strata fixed effects	No
Controls	Yes
Observations	905

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; EF refers to elaborated feedback; Strata fixed effects are excluded due to multicollinearity issues; Clustered standard errors at school level in parentheses; Wild bootstrap method controls for few (treated) clusters in data using bootstrapping to obtain critical values; Controls: baseline financial value, grade in mathematics, track, gender, language spoken at home, and time between tests; Note that data on teacher characteristics were only collected for a subsample of students.