WPS8161

Policy Research Working Paper 8161

Does Financial Education Impact Financial Literacy and Financial Behavior, and If So, When?

Tim Kaiser
Lukas Menkhoff

WORLD BANK GROUP
Development Economics Vice Presidency
Strategy and Operations Team
August 2017

Abstract

A meta-analysis of 126 impact evaluation studies finds that financial education significantly impacts financial behavior and, to an even larger extent, financial literacy. These results also hold for the subsample of randomized experiments (RCTs). However, intervention impacts are highly heterogeneous: financial education is less effective for low-income clients as well as in low- and lower-middle income economies. Specific behaviors, such as the handling of debt, are more difficult to influence and mandatory financial education tentatively appears to be less effective. Thus, intervention success depends crucially on increasing education intensity and offering financial education at a "teachable moment."

This paper is a product of the Strategy and Operations Team, Development Economics Vice Presidency. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org.
The authors may be contacted at lmenkhoff@diw.de.

Does Financial Education Impact Financial Literacy and Financial Behavior, and If So, When?

Tim Kaiser and Lukas Menkhoff

JEL classification: D 14 (personal finance), I 21 (analysis of education)
Key words: financial education, financial literacy, financial behavior, meta-analysis, meta-regression, impact evaluation

Tim Kaiser is a research associate at the University of Kiel, Germany and the German Institute for Economic Research (DIW Berlin); his email address is tkaiser@diw.de.

Lukas Menkhoff (corresponding author) is the head of department of International Economics at the German Institute for Economic Research (DIW Berlin) and Professor of Economics at the Humboldt-University of Berlin; his email address is lmenkhoff@diw.de.

Acknowledgements: We thank the authors who responded to our requests to provide their datasets or further details about their studies for their kind cooperation. Moreover, we appreciate valuable comments from participants at the Research in Behavioral Finance Conference 2016 in Amsterdam, the Meta-Analysis in Economics Research Network Colloquium 2016 in Conway, What Works Global Summit 2016 in London, the Conference in Behavioral Economics and Financial Literacy 2016 in Barcelona, and seminar participants in Berlin, Halle, Hamburg, Kampala, Kiel, and Vienna. In particular, we thank the editor (Eric Edmonds), three anonymous referees, Martin Brown, Nathan Fiala, Greg Fisher, Antonia Grohmann, Roy Kouwenberg, Jochen Kluve, Andreas Lutter, Christian Martin, Olivia Mitchell, Bob Reed, Anna Sokolova, Tom Stanley, Bertil Tungodden, Ludger Wössmann, and Dean Yang. Research assistance by Melanie Krüger and Iven Lützen, and financial support by DFG through CRC TRR 190 are gratefully acknowledged.

I. Introduction

The financial behavior of consumers and small-scale entrepreneurs is receiving increased interest. Evidence suggests a remarkable incidence of suboptimal individual financial decisions despite the fact that these decisions are highly relevant for individual welfare. The most prominent case of such an important financial decision in advanced economies is the amount and kind of retirement savings (cf. Duflo and Saez 2003). Studies show that undersaving is prevalent in many advanced economies and that households tend to save in inefficient ways, indicating that many may be unable to cope with the increasingly complex financial markets (e.g., Lusardi and Mitchell 2007; Choi et al. 2011; Behrman et al. 2012; van Rooij et al. 2012). This kind of behavior also stretches across other areas, including portfolio composition (Campbell 2006;Choi et al. 2010;Bucher-Koenen and Ziegelmeyer 2014;von Gaudecker 2015), excessive and overly expensive borrowing (Stango and Zinman 2009; Gathergood 2012; Agarwal and Mazumder 2013; Gerardi et al. 2013; Zinman 2015), as well as participation in financial markets in general (van Rooij et al. 2011). Related problems arise in developing countries often with even more serious consequences as people are exposed to heavy shocks without having sufficient insurance or mitigation instruments (e.g., Cole et al. 2011; Drexler et al. 2014; Gibson et al. 2014; Sayinzoga et al. 2016). All this strongly motivates providing financial education to foster financial behavior.

In surprising contrast to this obvious motivation for financial education stands the lack of compelling evidence that providing financial education is an effective policy for targeting individual financial behavior (Hastings et al. 2013; Zinman 2015). Narrative literature reviews are inconclusive, either emphasizing the effectiveness of education measures (e.g., Fox et al. 2005; Lusardi and Mitchell 2014) or emphasizing the opposite (e.g., Willis 2011). Further, the two available meta-analyses of this issue do not converge in their findings: Fernandes et al. (2014) summarize overall unreliable effects of financial education, whereas Miller et al. (2015) show that education can be effective in targeting specific financial behaviors. Given this inconclusive evidence on a most important issue, what can we learn in order to explain the heterogeneity in findings and to make financial education more effective?

We go beyond the extant literature and systematically code the circumstances of financial education for our meta-analysis. This allows us to examine the determinants of a positive impact of education. Another unique characteristic of our analysis is the focus on both objectives of financial education (i.e., improvements in financial literacy and financial behavior). Hence, we investigate the role of financial literacy for financial behavior in a unified setting. Finally, our study benefits from a rapidly rising field (see figure S1.1 in the supplemental appendix S1).

We follow the established procedures for the meta-analysis approach (e.g., Lipsey and Wilson 2001). The result is a sample of 126 studies reporting 539 effect sizes. Studies targeting entrepreneurs and exclusively measuring business outcomes (such as revenues) are omitted by design. We only

consider studies reporting about interventions, such as trainings and counseling efforts. Thus, we focus strictly on exogenous variation in financial education and neglect works exclusively analyzing the possible impact of cross-sectional (baseline) differences in financial literacy on financial behavior. Finally, we carefully code interventions as we examine in detail how financial education was delivered to the target groups.

Our meta-analysis results in six principle findings: (i) increasing *financial literacy* helps. Financial education has a strong positive impact on financial literacy with an effect size of 0.26 (i.e., above the threshold value of 0.20 that characterizes "small" statistical effect sizes [see Cohen 1977]). Moreover, effects on financial literacy are positively correlated with effects on financial behavior; (ii) financial education has a positive, measurable *impact on financial behavior* with an effect size of 0.09. An effect size of 0.08 is still found under rigorous randomized experiments (RCTs); (iii) effects of financial education depend on the *target group*. First, teaching low-income participants (relative to the country mean) and target groups in low- and lower-middle income economies has less impact, which is an obvious challenge for policymakers targeting the poor. Second, it appears to be challenging to impact financial behavior as country incomes and mean years of schooling increase, probably because high baseline levels of general education and financial literacy cause diminishing marginal returns to additional financial education; (iv) success of financial education depends on the *type of financial behavior* targeted. We provide evidence that borrowing behavior may be more difficult to impact than saving behavior by conventional financial education; (v) increasing *intensity* supports the effect of financial education; and (vi) the *characteristics* of financial education can make a difference. Making financial education mandatory is associated with deflated effect sizes. By contrast, a positive effect is associated with providing financial education at a "teachable moment" (i.e., when teaching is directly linked to decisions of immediate relevance to the target group (cf. Miller et al. 2015:13).

Complementing these findings, the meta-analysis also provides interesting non-results because several characteristics of financial education are without systematic impact on financial behavior. These include the age and gender of participants, the setting, or the choice of intervention channel through which financial education is delivered.

The findings reported above clearly motivate to implement financial education because it can positively affect financial literacy and financial behavior. However, its limited effectiveness raises two additional problems for policymakers: First, what can be done to make financial education generally more effective? Second, as a particularly obstinate aspect of the general question raised before, how can one reach those people who do not participate voluntarily? Problematic groups in this respect include low-income individuals, residents of low-income countries, and all those who do not self-select into education measures, as indicated by negative effects from mandatory courses and RCTs. For these groups, it appears that financial education needs an improved approach to be successful. More research

and experience is necessary to better identify the determinants of successful financial education (e.g., Hastings et al. 2013).

Our study follows several earlier survey studies about financial education. Most of these studies have a narrative character, among them widely cited works such as Fox et al. (2005), Willis (2011), Hastings et al. (2013), and Lusardi and Mitchell (2014). This gives the authors some flexibility about selecting and interpreting the most relevant studies. A quantitative meta-analysis is more rigid in approach but has the advantages that transparent rules of procedure ensure replicable results and that quantitative relations can be derived. Overall, narrative surveys and meta-analyses complement each other.

We perform a meta-analysis because there are just two earlier systematic accounts of the financial education literature that leave much room for more research. The study by Miller et al. (2015) covers only 19 papers due to its extremely restrictive selection criteria, requiring interventions on *identical* outcomes. This limits the sample sizes to about five studies and estimates per subsample, which does not allow investigating the sources of heterogeneity.

Thus, the most similar study to our work is Fernandes et al. (2014), which covers 90 effect sizes from financial education reported in 77 papers. Despite an overlap of 44 percent with their sample of studies, our research differs in four crucial ways, which explains our new results: (i) most important is that we analyze determinants of program effectiveness in a broader way by applying respective coding; (ii) we consider various outcomes per study (on average about four per study) and their respective effect sizes; moreover, (iii) we cover recent and mostly randomized experiments providing evidence of effective interventions; and (iv) we cover additional studies focusing exclusively on financial literacy as the outcome variable.

This paper is structured in seven further sections. Section 2 introduces our meta-analytic approach. Section 3 describes our data. Section 4 provides first results of the meta-analysis, while section 5 uses these results to explain heterogeneity of financial education treatment effects. Robustness tests are mentioned in section 6, and section 7 concludes with policy considerations and venues for future research.

II. Meta-Analytic Method

Meta-analysis is a quantitative method to synthesize findings from multiple empirical studies on the same empirical research question. In a meta-analysis, the dependent variable is comprised of a summary statistics reported in the primary research reports, while the explanatory variables may include characteristics of the research design, the sample studied, or, in case of impact evaluations, the policy intervention itself (cf. Stanley 2001: 131). Meta-analyses can provide answers to two specific

questions (cf. Muller 2015; Pritchett and Sandefur 2015; Vivalt 2015). First, is the combined (statistical) effect across all studies reporting effects of similar interventions on similar outcomes significantly different from zero? And, second, what explains heterogeneity in the reported findings?

In order to be able to aggregate summary statistics reported across heterogeneous studies, one must standardize these statistics into a common metric. If all studies would operationalize and measure outcomes in the same unit, meta-analysis could be performed directly using *economic* effect sizes (e.g., elasticities or marginal effects) in contrast to *statistical* effect sizes (cf. Stanley and Doucouliagos 2012: 23). This, however, is rarely the case in a large sample of heterogeneous (quasi-) experimental impact evaluations.

Thus, we use a standard approach of coding a variable capturing intervention success and impact. Our impact measure (effect size) is the standardized mean difference (SMD) for each treatment effect estimate. We use the bias corrected standardized mean difference (Hedges' g) as our effect size measure, which is defined as the mean difference in outcomes between the treatment (M_T) and control (M_C) (i.e., the treatment effect) groups as a proportion of the pooled standard deviation (SD_p) of the dependent variable:

$$(1) \qquad g = \frac{\mathrm{M}_T - \mathrm{M}_\mathrm{C}}{\mathrm{SD}_\mathrm{p}}$$

with

$$(2) \qquad SD_p = \sqrt{\frac{(\mathrm{n}_\mathrm{T}-1)\, SD_T{}^2 + (\mathrm{n}_\mathrm{C}-1)\, SD_C{}^2}{n_T{}^2 + n_C{}^2 - 2}}.$$

where n_T and SD_T are the sample size and standard deviation of the treatment group, and n_C and SD_C are for the control group. Additionally, we capture the standard error of each standardized mean difference (g), which is defined as:

$$(3) \qquad SE_g = \sqrt{\frac{\mathrm{n}_\mathrm{T}+\mathrm{n}_\mathrm{C}}{\mathrm{n}_\mathrm{T}\mathrm{n}_\mathrm{C}} + \frac{g^2}{2(\mathrm{n}_\mathrm{T}+\mathrm{n}_\mathrm{C})}}$$

Hedges' g informs about the size and direction of an effect in scale-free standard deviation units. This metric is only slightly different from other popular effect size measures in experimental impact evaluations, such as Cohen's d and Glass Δ (see, e.g., Banerjee et al. 2015). Hedges' g, however, introduces minor corrections that reduce bias in the effect size estimate in cases with small sample sizes and when the sample sizes of treatment and control groups are unequally distributed. Results are qualitatively robust to using alternative measures or relying on (partial) correlations (cf. Lipsey and Wilson 2001).

As a rule of thumb, Cohen (1977) suggests that effect sizes smaller than 0.20 should be considered as a "small effect"; effect sizes around 0.50 indicate a "medium effect"; while effect sizes greater than 0.80 constitute "large effects." Where pure mean comparisons, standard deviations, and sample sizes for each experimental outcome are not reported directly we exhaust all possibilities to calculate or estimate effect sizes (g) and its corresponding standard error from the range of available statistical data (cf. Lipsey and Wilson 2001).

In the estimation of summary effects of the literature, our main approach follows a full pooling least squares meta-regression framework (e.g., Card et al. 2015). Accordingly, the financial education treatment effect (g) can be explained by exogenous, observable characteristics, the impact g on an outcome *i,* reported in study *j* is expressed as a linear function

$$(4) \qquad g_{ij} = \alpha + x_{ij}\beta + \epsilon_{ji}$$

where $x_{ij}\beta$ is a vector of observable (exogenous) study-level covariates, such as intensity of intervention, α is an intercept, and ϵ_{ji} denotes an error-term independent from $x_{ij}\beta$. We estimate our models using multiple effect sizes per study and account for heteroscedasticity by clustering standard errors at the study-level. Reassuringly, results are not sensitive to a set of changes in estimation strategy and accounting for publication selection bias (see section 6 and supplemental appendix S3).

III. Sample Description

This section describes the selection of studies, the extraction of effect sizes and study-level covariates, and types of financial education programs.

Selection of Studies

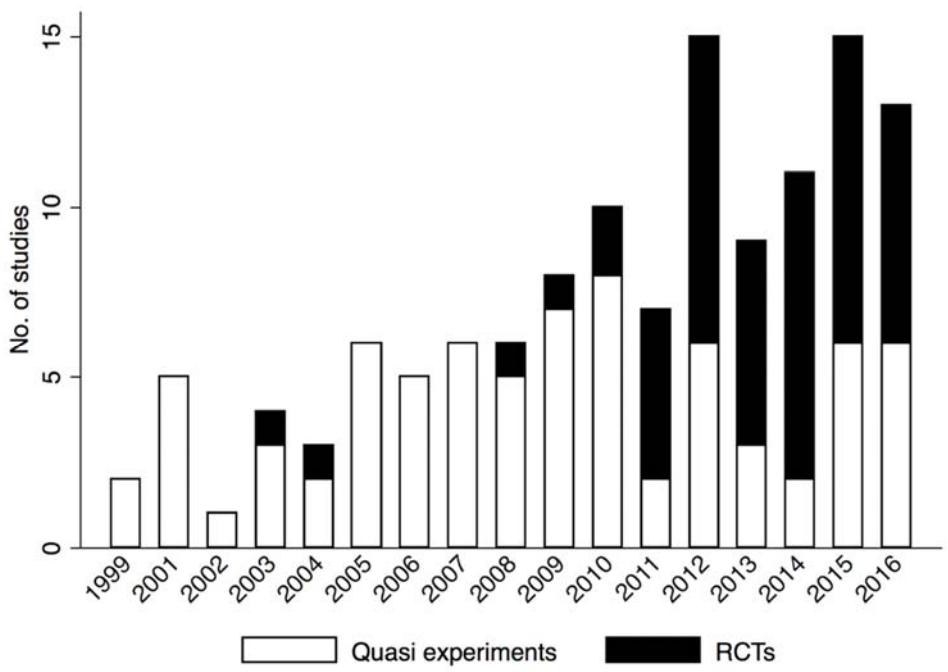
We follow the established meta-analytical protocol (cf. Lipsey and Wilson 2001: 23, Stanley 2001: 143). This starts with systematically searching the relevant databases, including working papers, for the following keywords: (i) financial literacy; (ii) financial knowledge; (iii) financial education; (iv) financial capability; and (v) combinations of these keywords with "intervention." Moreover, we consider all records from meta-analyses (Fernandes et al. 2014; Miller et al. 2015) and narrative literature reviews (Fox et al. 2005; Collins and O'Rourke 2010; Willis 2011; Xu and Zia 2012; Hastings et al. 2013; Blue et al. 2014; Lusardi and Mitchell 2014). This search resulted in over 500 potentially relevant published journal articles and over 600 results from working paper databases with some apparent overlap. We stopped collecting studies in October 2016 (see appendix S1).

From this collection, we drop studies that do not meet our three criteria for inclusion: (i) reporting on impacts of an exogenous educational intervention on financial literacy and/or financial behavior;

(ii) providing a quantitative assessment of intervention impact that allows coding an effect size statistic (g) and its standard error; and (iii) relying on an observed counterfactual in the estimation of intervention impacts. This selection process leads to a final sample of 126 independent intervention studies that report 539 effect sizes (further details in tables S1.1 and S1.2 in the supplemental appendix S1). Of these, 90 studies report 349 effect sizes on financial behavior, and 67 studies report 190 effect sizes on financial literacy. Among these 90 plus 67 studies, there are 31 studies reporting effect sizes on both financial literacy and behavior.

RCTs are rare in the early years of the literature, but their share has risen dramatically, with the majority of studies conducted from 2011 onward being randomized evaluations (see figure 1). This development in the literature is very favorable for meta-analyses, since it ensures a high internal validity of research findings reported in the primary studies and helps to clearly distinguish between selection and treatment effects.

Figure 1. Number of Studies in Our Sample by Research Design per Year



Source: Authors' calculations based on the data source discussed in the text.

Extraction of Effect Size Estimates and Study Descriptors

As the next step, we code the effect of financial education on *financial literacy* (i.e., a measure of performance on a financial knowledge test), since knowledge development is the primary goal of financial education (Hastings et al. 2013; Lusardi and Mitchell 2014). Moreover, we code treatment effects of financial education on several *financial behaviors* (see table S1.2 in the supplemental appendix S1), such as an increase in savings after the treatment. Multiple estimates per study are considered if multiple outcomes, time-points, or treatments are reported; however, results are robust to aggregating all effects per study into one synthetic effect size. Further details about this process are described in supplemental appendix S1.

Types of Financial Education Programs

Our dataset includes four main types of financial education programs. First, and most frequent, are evaluations of *classroom financial education* (approximately 83 percent of all estimates) in various settings, such as schools, universities, the workplace, or specific sites such as savings groups or microfinance institutions. These studies are quasi-experiments or RCTs, in which the researcher has control over content, intensity, and survey design in order to measure specific outcomes. There is an increasing interest in the literature in multiple-treatment and cross-over designs to investigate optimal delivery strategies and potential causal mechanisms (i.e., Drexler et al. 2014; Carpena et al. 2015; Skimmyhorn 2016). These studies have high internal validity but may report site-specific effects that causally interact with unobserved features of the specific sites (cf. Muller 2015). Additionally, measurement of outcomes is typically in the short or medium run (approx. 65 percent), since long time series are usually not available. A different strand of the literature evaluating this type of program looks at classroom financial education utilizing (plausibly exogenous) variation in (mandatory) school financial education mandates (e.g., Tennyson and Nguyen 2001; Brown et al. 2016). These studies are typically quasi-experimental in nature, and, while possibly weaker in internal validity, possess high external validity, since they typically have large sample sizes and measure relatively long-run effects on behavioral outcomes, such as savings.

A second type of intervention is *online financial education* (approx. 8 percent of estimates). While similar in research design to experiments on classroom financial education, these studies usually estimate the effect of certain online modules on financial literacy and behavior and typically evaluate instructional videos or interactive applications.

The third type of financial education treatments evaluated in the literature are *individualized counseling interventions* (two percent of estimates). These have been mainly studied in the US and typically study outcomes related to the handling of (mortgage) debt.

As a fourth and last type, we identify *informational and behavioral nudges*, such as information fairs at the workplace and informational brochures (seven percent of estimates). These studies typically evaluate behavioral change in response to these low-intensity treatments. There is one study in our sample that studies the effect of a behavioral nudge in the form of "financial edutainment" in mass-media (cf. Berg and Zia 2013). This is an intervention designed to impact financial behaviors through a non-cognitive channel (as opposed to increasing financial knowledge), and the included study evaluates the impact of financial messages inserted into episodes of a popular television series in South Africa.

IV. Results from Meta-Analysis

We report the mean effects for all studies (section 4.1) and then for subsamples: financial literacy and financial behavior (section 4.2), types of financial education programs (section 4.3), research designs (section 4.4), and different country groups (section 4.5).

Summary Effects of Financial Education

Here we discuss the average effects of financial education on financial literacy and financial behavior. Based thereon, we study the relation between these two outcomes. As a starting point, we note that the summary effect of financial education on all kinds of reported outcomes is estimated to be $g = 0.148$ ($p = .000$, $n = 539$). However, heterogeneity in effect sizes is high, indicating that outcomes could be disaggregated for meaningful analyses.

Financial behavior. We find that the average impact of educational interventions on financial behaviors is statistically highly significant ($g = 0.086$) (see table S1.3 in the supplemental appendix S1). The main reason that we get a more favorable result than Fernandes et al. (2014) is that we profit from a moderate, positive time trend (more details in supplemental appendix S2). To compare the magnitude of this effect size to results from health promotion on behavioral change (e.g., weight loss and nutrition in obesity studies), Portnoy et al. (2008) report in their meta-analysis of 75 RCTs an average effect size of about 0.1.

*Financial literac*y**.** The average impact of financial education on financial literacy is substantially higher ($g = 0.263$, $p = .000$, $n = 190$) than the one on financial behavior (see figure S1.2 and table S1.3 in the supplemental appendix S1). Moreover, financial education explains 1.7 percent of the variance in financial knowledge and, thus, appears to be only slightly less effective than educational interventions in other domains such as math and science instruction (cf. Fernandes et al. 2014: 1867). To put this effect size in perspective: the meta-analysis of 225 studies by Freeman et al. (2014) reports an average effect size of around 0.47 for studies evaluating student performance in response to

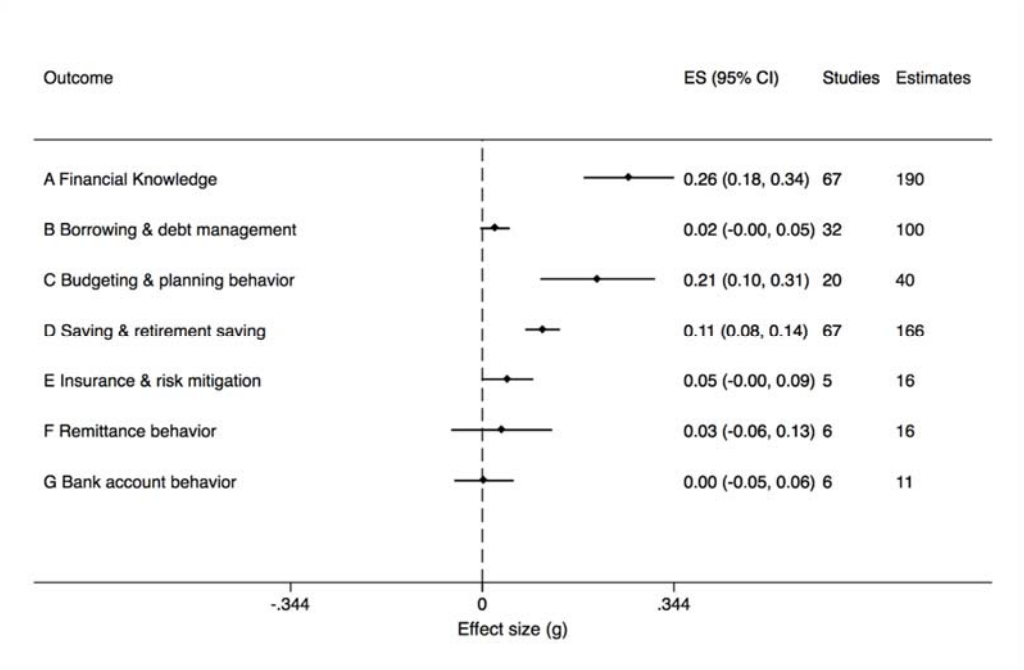
alternatives to lecturing in undergraduate science education; however, these interventions occur in a university context and last for a full semester.

Relationship between financial literacy and behavior. The intuition is that increases in financial literacy scores are an important intermediate result in a causal chain expected to lead to behavior change (e.g., Grohmann et al. 2015; Fort et al. 2016). Indeed, for a sample of 31 studies, we find in a regression with standard errors clustered at the study-level that the effect size on financial literacy is a statistically significant predictor of effect size on financial behavior ($b = 0.230$, $p = .022$). Thus, an increase of one standard deviation unit in financial literacy scores is related to an average increase of 0.23 standard deviation units of the financial behaviors studied. However, the non-overlapping confidence intervals of these effect sizes also indicate that these two elements of the causal chain should be analyzed separately when attempting to explain the heterogeneity in effect sizes.

Effect Sizes by Type of Financial Behavior

Figure 2 shows the average effect size for the seven categories of financial behaviors targeted by the educational interventions in our sample.

Figure 2. Forest Plot of Effect Sizes by Type of Financial Behavior Studied



Source: Authors' calculations based on the data source discussed in the text.

Average effect sizes for three out of seven categories of outcomes are clearly positive and highly statistically significant at the one percent level. Additionally, all confidence intervals for the different types of financial behaviors overlap each other, indicating that there are no extreme differences in impacts depending on the specific form of financial behavior targeted. In detail: (i) the average effect size on "budgeting" appears to be higher than those on downstream behaviors; and (ii) effect sizes related to saving and retirement saving appear to be higher than the average effect size of financial education on borrowing behavior; (iii) this latter average effect size is small ($g = 0.02$) and insignificant from zero; (iv) similarly, the average effect sizes for "insurance" ($g = 0.05$), "remittances" ($g = 0.03$), and "bank account behavior" ($g = 0.00$) are estimated to be small and insignificant from zero, although based on a few studies per category only. Thus, debt-related financial behaviors may be the most challenging to target through financial education (see Miller et al. 2015: 238). Overall, these findings correspond to the results provided by Fernandes et al. (2014) and Miller et al. (2015) and extend to our much larger sample.

Effect Sizes by Type of Financial Education Intervention

We form subsamples by the main types of financial education interventions, as discussed in section 3.3. First, we compare classroom financial education to three types of non-classroom delivery channels (online financial education, counseling, and informational/behavioral nudges). Second, we distinguish between financial education at school and two non-school settings (workplace and other settings). Panel A of table 1 shows results split by outcomes on financial literacy and financial behavior. While in-person classroom trainings appear to be (unconditionally) more effective than non-classroom delivery channels in increasing financial knowledge, we observe no statistically significant difference regarding impacts on financial behavior. Turning to the intervention setting, it appears that interventions in schools are more effective at increasing financial literacy but yield marginally significant smaller treatment effects on financial behavior. However, we note that these relations are obviously partially confounded with several other relevant variables (e.g., the age of the participants, the delay in measurement, and research design), which indicates the importance of an examination in a multivariate setting (cf. section 5).

Table 1. Effect Sizes of Financial Education by Intervention Type, Research Design, and Country Groups

Outcome	Type	Studies	Obs.	ES (g)	SE_g	p-value	Diff. (t-value)
A Effect sizes by intervention channel & setting							
Fin. literacy	Classroom	58	135	0.294	0.054	0.000	0.106**
	Non-classroom	9	55	0.188	0.039	0.001	(2.015)
	- Online	*5*	*41*	*0.217*	*0.060*	*0.018*	
	- Counseling	*0*					
	- Nudge	*4*	*14*	*0.103*	*0.045*	*0.108*	
Fin. behavior	Classroom	70	317	0.084	0.013	0.000	-0.014
	Non-classroom	20	32	0.098	0.020	0.000	(0.452)
	- Online	*11*	*18*	*0.085*	*0.034*	*0.031*	
	- Counseling	*7*	*8*	*0.095*	*0.030*	*0.020*	
	- Nudge	*2*	*6*	*0.140*	*0.007*	*0.031*	
Fin. literacy	School	35	62	0.373	0.076	0.000	0.163***
	Non-school	32	128	0.210	0.035	0.000	(3.273)
	- Workplace	*1*	*1*	*0.164*	*0.063*		
	- Other	*31*	*127*	*0 210*	*0.035*	*0.000*	
Fin. behavior	School	27	90	0.057	0.014	0.000	-0.039*
	Non-school	63	259	0.096	0.014	0.000	(1.96)
	- Workplace	*17*	*47*	*0.121*	*0.049*	*0.023*	
	- Other	*46*	*212*	*0.090*	*0.015*	*0.000*	
B Effect sizes by research design							
Fin. literacy	RCTs	33	135	0.209	0.033	0.000	-0.185***
	Quasi-exp.	34	55	0.394	0.083	0.000	(-3.638)
Fin. behavior	RCTs	40	227	0.081	0.015	0.000	-0.012
	Quasi-exp.	50	122	0.093	0.022	0.000	(-0.661)
C Effect sizes by country group							
Fin. literacy	High income	53	123	0.328	0.058	0.000	0.183***
	Developing	14	67	0.145	0.031	0.000	(3.787)
	- Low	*3*	*6*	*0.219*	*0.069*	*0.086*	
	- Lower-middle	*6*	*44*	*0.155*	*0.047*	*0.023*	
	- Upper-middle	*5*	*17*	*0.092*	*0.023*	*0.017*	
Fin. behavior	High income	66	168	0.071	0.019	0.000	-0.027
	Developing	24	181	0.098	0.014	0.000	(-1.512)
	- Low	*6*	*39*	*0.161*	*0.038*	*0.009*	
	- Lower-middle	*12*	*90*	*0.091*	*0.008*	*0.000*	
	- Upper-middle	*6*	*52*	*0.06*	*0.023*	*0.045*	

Notes: Average effect sizes *(g)* estimated via OLS regressions of effect sizes fitting only an intercept. Sample is split by an indicator of intervention type, research design, or country group. "Channel" is a categorical variable operationalized in the form of four dummy variables: Classroom, Counseling, Online, and "Nudge" where "Nudge" is the default (omitted) category in the regressions. "Setting" is a categorical variable operationalized through three dummy variables: School, Workplace and Other where Other is the omitted category in the meta-regression analyses. Country groups are based on the World Bank Atlas method and refer to 2015 data on GNI per capita. Low-income economies are defined as those with a GNI per capita of $1,025 or less in 2015, lower-middle income economies are defined by a GNI per capita between $1,026 and $4,035, upper-middle income economies are those with a GNI per capita between $4,036 and $12,475, and high-income economies are defined by a GNI per capita greater than $12,475. Standard errors are clustered at the study level. ***, ** and * denote significance at the one percent, five percent and ten percent level.
Source: Authors' analysis based on data sources discussed in the text.

Effect Sizes by Research Design

Regarding research design, Fernandes et al. (2014: 1865) find that weaker research designs lead to inflated effect sizes. Thus panel B of table 1 compares average effect sizes as a function of research design. When we focus on financial behaviors as outcomes, RCTs show statistically highly significant (unconditional) effect sizes of 0.081. These are only slightly smaller than for quasi-experiments with 0.093, indicating that the small but positive significant effects of financial education exist even under the most rigorous empirical standards. RCTs also provide a significant positive effect of financial education on financial literacy with 0.209. Here the difference to other designs (effect size of 0.394) is significant at the one percent level.

Effect Sizes by Country Groups

To investigate another potential source of heterogeneity, we disaggregate our data by country groups. Panel C of table 1 shows effect sizes by country groups as classified by the World Bank based on 2015 GNI per capita. We find that effect sizes on *financial literacy* are significantly higher in developed (high income) economies (g = 0.328) than in developing economies (low income, lower- and upper-middle income economies, g = 0.145). Turning to effect sizes on *financial behavior*, this difference is statistically insignificant in this unconditional comparison but differences between country groups become more nuanced and statistically significant when controlling for other relevant variables (see section 5.2).

V. Explaining Heterogeneity in Financial Education Treatment Effects

Section 4 shows that the average effect size of financial education is accompanied by large heterogeneity. Thus, we examine whether there are factors explaining this heterogeneity. This will also suggest directions that future financial education policies might take in order to increase their impact on financial behavior.

Potential Correlates of Effect Size

The effectiveness of financial education is potentially influenced by the peculiarities of the specific intervention. Based on prior literature, we group these characteristics into four categories: (i) the research design; (ii) the intensity of education; (iii) the target group of education; and (iv) the characteristics of the education program.

(i) Regarding the *research design* of a financial education study, we expect the method of investigation (i.e., RCT vs. less rigorous designs) to be relevant. Second, the concrete measurement of an effect will influence the estimated size of impact. It is known that focusing on treatment on the

treated (TOT) (i.e., measuring a treatment effect on the population who actually *received or attended* the treatment) generally results in higher effect sizes than focusing on the intention to treat (ITT) effect (i.e., the population who was in principle *assigned* to treatment). However, ITT may be more relevant for policy (cf. Imbens and Wooldridge 2009: 15; Gertler et al. 2011: 73). Third, the delay between financial education treatment and measurement of the effect may negatively influence the effect size since effects of the intervention may decay over time (cf. Fernandes et al. 2014: 1867). Additionally, we control for the precision of effect size estimation by the inverse standard error (or the [squared] standard error, see supplemental appendix S3). All these variables are defined in table 2, which also provides descriptive statistics.

(ii) A core variable of financial education interventions is the *intensity of education* (i.e., the number of hours taught). It is expected that higher intensity will support the effect. However, the time-frame over which the financial education intervention is delivered to the target group may also be of importance. We expect differences between high intensity and low intensity relative to the duration. Thus, we code the *hours of financial education per week* (i.e., intensity per week) and the *duration* of the intervention in weeks to investigate this issue.

(iii) The expectation regarding a possible relation between the *target group* of education and effectiveness of financial education is as follows. Generally, learning is easier for younger people, younger people may be more open to new concepts and their baseline financial literacy scores are low (e.g., Lusardi and Mitchell 2014), meaning that the *age* of the target group may be negatively related to the effect size of financial education. Second, a gender gap in financial literacy is treated as a stylized fact in the literature (cf. Lusardi and Mitchell 2014) which may also translate to gender differences in treatment effects. Thus we include the *percentage of women* in the sample. Third, it is expected that the acquaintance of the target group with an educational environment may be helpful. As a proxy for such openness to education, we take the *income* of the target group relative to the overall population. Fourth, we expect that the overall institutional level of education should support domain-specific educational efforts (Jappelli 2010). As a proxy for this potential relationship, we take a country's population mean *years of schooling* as reported by the United Nations Development Program Human Development Reports. Additionally, we augment our data with *country-level financial literacy* data from a 2015 global financial literacy survey (Klapper et al. 2015). We hypothesize that financial education interventions may yield higher effects when the population baseline financial literacy is lower, indicating more room for improvement through education. Finally, as a control variable we code the country of intervention according to the World Bank country group classifications.

(iv) Regarding the *characteristics of the education program*, it seems interesting whether the *channel* (i.e., classroom, online, individual counseling, etc.) is important in explaining education effectiveness, since these formats come with different trainer to participant ratios and may rely on

different pedagogical approaches to financial education. It may be that willingness to learn and change financial behavior is lower when financial education is *mandatory* (cf. Collins 2013) or motivation to participate in financial education is not intrinsic but driven by *incentives* provided by the offering institution. Lastly, these characteristics may be correlated with specific *settings* (i.e., at school or at the workplace).

Next, and going further in this direction, it is coded whether participants are educated at a *teachable moment* (i.e., that they have the possibility to apply their knowledge in a concrete case of interest to them, e.g., Doi et al. 2014; Miller et al. 2015). Thus, we capture whether the education addressed immediate financial issues (such as borrowers already in default, or migrants confronted with deciding through which channel remittances are sent). Alternatively, financial education was generic and offered at an unspecific moment, as is often the case in large scale financial education programs (e.g., Bruhn et al. 2014).

Table 2. Summary Statistics

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
A Descriptive statistics at the study-level					
RCT	126	0.405	0.493	0.000	1.000
TOT	115	0.452	0.500	0.000	1.000
Delay	93	82.231	273.613	0.000	1566
1/SE	126	57.535	210.450	2.480	1636.712
Intensity	87	11.211	14.929	0.100	87.000
Duration	76	7.341	14.150	1.000	103.000
Age	109	30.717	14.120	9.000	63.870
Percent female	123	54.011	18.493	0.000	100.000
Low income clients	102	0.529	0.502	0.000	1.000
Years of schooling	126	11.270	2.843	3.200	13.600
FL in population	124	50.419	11.658	24.000	66.000
Mandatory	96	0.292	0.457	0.000	1.000
Incentivized	86	0.314	0.467	0.000	1.000
Teachable moment	126	0.397	0.491	0.000	1.000
B Descriptive statistics at the estimate-level					
RCT	539	0.672	0.470	0.000	1.000
TOT	510	0.282	0.451	0.000	1.000
Delay	463	93.742	292.025	0.000	1566.000
1/SE	539	41.260	124.389	2.740	957.167
Intensity	451	15.384	23.444	0.100	144.000
Duration	434	7.908	14.236	1.000	103.000
Age	494	31.814	11.720	9.000	63.870

-continued-

Percent female	525	52.923	18.200	0.000	100.000
Low income clients	451	0.681	0.467	0.000	1.000
Years of schooling	539	9.890	3.463	3.200	13.600
FL in population	523	44.170	14.668	24.000	66.000
Mandatory	480	0.240	0.427	0.000	1.000
Incentivized	445	0.247	0.432	0.000	1.000
Teachable moment	539	0.479	0.500	0.000	1.000

Notes: "RCT" is a dummy variable with "1" if selection into treatment was conducted through randomization and "0" otherwise (such as matched designs). "TOT" is a dummy variable with "1" if the effect size estimate is derived from the treatment effect on the treated and "0" if it is derived from the ITT estimate. "Delay" is a continuous variable indicating the delay between treatment and measurement of outcomes in weeks. "1/SE" is the inverse standard error for each effect size estimate. "Intensity" is the total number of hours of financial education exposure to the treated. "Duration" indicated the time-frame of financial education in weeks. "Age" is the mean age of the sample in years. "Percent Female" is the relative frequency of female participants in the sample in percent. "Low income" is a dummy variable with "1" if the mean annual income per capita of the sample is below the country average income per capita. "Mandatory" is a dummy variable with "1" indicating mandatory participation in financial education and "0" voluntary participation. "Incentivized" is a dummy variable with "1" when incentives to participate where provided and "0" if participation was unconditional on incentives. "Teachable moment" is a dummy variable indicating whether the financial education intervention was offered at a teachable moment.
Source: Authors' analysis based on data sources discussed in the text.

Meta-regression Models Explaining Intervention Impacts

This section examines determinants of financial education effectiveness using a multivariate meta-regression framework including the above discussed potential correlates as right-hand side variables. Our procedure is motivated by economic and econometric considerations. From an economic point of view, we aim for including all variables that have a substantial theoretical foundation. From an econometric viewpoint, the specification should be parsimonious, especially in the presence of a relatively small sample size of studies.

Thus, we start with a specification where we include all reasonable and available variables (table 3, column 1). In order to keep the number of studies considered high, we impute average or default values for missing observations (we show in supplemental appendix S3 that our main results are insensitive to imputation). The discussion considers groups of variables in four blocks, following their introduction in section 5.1.

Table 3. Explaining Heterogeneity in Effect Sizes on Financial Behavior

	(1) All	(2) All	(3) RCTs	(4) Low inc. econ	(5) High / middle inc. econ	(6) Low income clients
RCT	-0.070**	-0.068**		-0.209**	-0.079**	-0.066**
	(0.027)	(0.028)		(0.091)	(0.036)	(0.032)
TOT	0.079***	0.068**	0.012	-0.016	0.076**	0.031
	(0.027)	(0.027)	(0.040)	(0.066)	(0.035)	(0.032)
Delay	0.000	0.000	-0.001**	-0.001**	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
1/SE	-0.000	-0.000	0.000	-0.003	-0.000	0.000
	(0.000)	(0.000)	(0.001)	(0.002)	(0.000)	(0.000)
Intensity / week	0.004**	0.004***	0.007***	0.004**	0.003	0.004***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)	(0.001)
Duration	-0.000	-0.000	-0.001	-0.001	-0.000	0.000
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
Age	-0.001					
	(0.001)					
Percent female	-0.000					
	(0.001)					
Low income clients	-0.065***	-0.055***	-0.074***	-0.042**	-0.048**	
	(0.020)	(0.017)	(0.024)	(0.019)	(0.021)	
Years of schooling	-0.016***	-0.019***	-0.016**	-0.026***	-0.025***	-0.011*
	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.006)
FL in population	-0.003					
	(0.002)					
Country group						
a) Low/lower-mid. inc. econ.	-0.129*	-0.093**	-0.092**			-0.059
	(0.073)	(0.036)	(0.042)			(0.042)
b) Upper-mid. inc. econ.	0.000					
	(0.060)					
Channel						
a) Classroom	-0.003					
	(0.028)					
b) Counseling	-0.018					
	(0.033)					
c) Online	-0.028					
	(0.028)					
Setting						
a) School	0.022					
	(0.023)					
b) Workplace	0.041					
	(0.036)					
Mandatory	-0.074***	-0.051**	-0.078*	-0.015	-0.065**	-0.052
	(0.024)	(0.023)	(0.044)	(0.042)	(0.025)	(0.033)
Incentivized	-0.012					
	(0.029)					
Teachable moment	0.079***	0.064**	0.016	0.025	0.069**	0.072**
	(0.021)	(0.026)	(0.035)	(0.026)	(0.029)	(0.032)
Constant	0.477***	0.332***	0.338***	0.514***	0.406***	0.188*
	(0.157)	(0.079)	(0.095)	(0.110)	(0.114)	(0.095)
R^2	0.210	0.183	0.149	0.170	0.204	0.109
n (Studies)	90	90	40	18	72	44
n (Effect sizes)	349	349	227	129	220	234

Notes: Non-standardized coefficients from OLS regressions. Dependent variable in columns (1) and (2) is effect size (Hedges' g) on financial behavior in the full sample of studies reporting on financial behavior as an outcome. Column (3) shows results for RCTs only. Column (4) and (5) show results for financial behavior split by country groups. Column (6) limits the sample to classroom trainings only. Robust standard errors clustered at the study-level in parentheses. ***, ** and * denote significance at the one percent, five percent and ten percent level.
Source: Authors' analysis based on data sources discussed in the text.

Research design. Starting with the research design of the underlying primary studies, we find that RCTs report—*ceteris paribus*—slightly smaller effect sizes than non-RCTs, which is in line with earlier presumptions (see table 1, panel B). However, now this difference is statistically significant (see column 1 of table 3). As expected, the operationalization of treatment effects as TOT-estimates leads to inflated effect size estimates. Apparently, the delay between intervention and measurement of outcomes does not seem to be systematically related to effect sizes in this estimation (cf. supplemental appendix S3 for an alternative approach and investigation of heterogeneous treatment effects depending on delay in measurement). In addition, estimates with large inverse standard errors are associated with smaller effect sizes, indicating that larger and more precise studies report smaller effect sizes overall. However, this coefficient is small in size and insignificant.

Intensity. Turning to the relationship between intensity per week and duration, column 1 of table 3 shows that intensity has a significant positive effect on treatment effects on financial behavior. Thus, an increase of one hour of financial education per week leads to a 0.004 standard deviation unit increase in the impact on financial behaviors studied. Considering that the average weekly duration is in this subsample is roughly nine weeks and weekly intensity is only about four hours, doubling the weekly intensity to eight hours while keeping everything else constant at the mean, would lead to an average treatment effect around 14 percent higher than the empirical mean predicted treatment effect in this fully specified model.

Target group. Among participant characteristics, age and gender are not significant explanatory variables. However, the coefficient on "low income clients" is highly significant and negative, indicating that these individuals are more difficult to educate. Regarding increasing mean years of schooling at the country level, returns to additional financial education appear to diminish. This is in line with results from two studies in very different contexts (Europe and India) that report higher treatment effects for lower-educated individuals and diminishing returns to financial education upon higher baseline levels of education (cf. Cole et al. 2011; Fort et al. 2016). Similarly, the coefficient for baseline financial literacy in the population is also negative, albeit statistically insignificant. While these results suggest declining marginal returns to financial education, the negative effect for low- and lower-middle income economies—and also the above-mentioned coefficient on low-income clients—shows a countervailing influence from challenging groups or country circumstances.

Characteristics of education. Regarding the channel variables, column 1 shows that no alternative channel appears to be generally more or less effective than financial education in classroom settings or informational nudges (omitted category). The same is true for the setting of the intervention where school and workplace settings are not systematically different from other settings. However, mandatory financial education and implementing financial education at a "teachable moment" appear to be important. Specifically, we find, that making financial education mandatory decreases effect sizes

by 0.074 standard deviation units: The predicted value for effect size on financial behavior in mandatory formats with everything else kept equal at the (empirical) mean would be only $g = 0.030$ ($SE = 0.020$, $p = .134$); thus, economically small and statistically insignificant from zero. In contrast, offering financial education at a teachable moment increases effect sizes by 0.079 standard deviation units. Thus, the predicted value for effect size on financial behavior would be *ceteris paribus* $g = 0.124$ ($SE = 0.014$, $p = .000$) (i.e., statistically highly significant), roughly 48 percent larger than the unconditional average effect size found in the sampleand about 45 percent larger conditional on the empirical means for all other covariates in this full model.

Parsimonious specification. We reduce the above discussed fully specified model by keeping the variables on research design and intensity but otherwise eliminating the insignificant variables. Column 2 of table 3 describes the resulting reduced model that confirms the fully specified regression results from column 1. There are just some smaller changes in the estimated standard errors that occur at a few variables. This indicates that it is justified to rely on the parsimonious specification, in particular when we analyze subsamples with a much smaller number of observations in the following.

Meta-Regression Models for Subsamples

Given the large degree of heterogeneity across the 90 studies and their underlying financial education programs, we move to an analysis of more homogenous subsamples.

RCTs only. Many will agree that RCTs fulfill the most rigorous requirements, implying that results limited to this subsample of studies are indeed reliable. We do not prefer this procedure because many observations are lost. Nevertheless, it is reassuring that results qualitatively hold, as shown in column 3 of table 3 for the subsample of 40 RCTs covering 227 effect sizes. However, while the negative coefficient for mandatory courses remains to be large in magnitude and statistically (marginally) significant, the coefficient for teachable moment loses explanatory power in this estimation.

Interventions in low and lower-middle income economies. This subsample covers 18 studies that report 129 effect sizes (see column 4 of table 3). Again, all coefficients have the same sign and similar magnitude as in our parsimonious specification (column 2 in table 3), but differences in standard errors arise. While intensity of the intervention remains a strong predictor and low-income clients in low-income economies also benefit significantly less from financial education, mandatory formats and timing in the sense of offering financial education at a teachable moment appear less predictive of treatment effects.

Interventions in upper-middle and high-income economies. Turning to the 72 studies that examine financial education in more affluent economies (column 5 of table 3), we find that results again are qualitatively very similar to the pooled analysis in column 2. Here, the opposing coefficients for mandatory formats and offering financial education at a teachable moment are statistically significant

at the five percent -level, indicating that these effects may be primarily driven by interventions in middle or high income economies.

Interventions for low-income individuals. Examining the subsample of 44 studies focusing on low-income individuals results in a similar picture arising. Effects appear to be higher with increased training intensity and offering financial education at a teachable moment. However, country-level years of schooling and country income are now only marginally significant and insignificant covariates, respectively. Additionally, the coefficient for mandatory courses still has the same sign and similar magnitude, but is estimated with a larger standard error.

*Disaggregating financial behaviors and financial behaviors by target group***.** As discussed in section 4.2, it appears to be easier to affect financial behaviors in terms of (retirement-) savings and budgeting compared to borrowing behavior. Thus, we disaggregate the sample into three categories of financial behaviors and search for potentially heterogeneous effects of our main explanatory variables. We reduce the choice of variables for some subsamples to avoid problems with degrees of freedom due to relative few observations.

Column 1 of table 4 shows results for the subsample of 32 studies reporting effect sizes on borrowing behavior. This result matches our main results of the aggregated sample of effect sizes (column 2 of table 3) with significant positive effects from increased intensity, negative effects for low-income target groups, and countries, negative effects from making financial education mandatory and positive effects from offering financial education at a teachable moment. Column 2 of table 4 shows results for the subsample of 20 studies that focus on borrowing as the outcome and have low-income clients as the target group. Again, results are nearly identical. However, the delay in measurement is now a marginally significant predictor: effect sizes in this sample seem to diminish as time between intervention and measurement of outcomes increases. Hence, treatment effects on debt related behaviors among low-income individuals may be shorter-lived.

Turning to effect sizes reported in 67 studies on (retirement-) saving (column 3 of table 4), we observe that the relevant variables from our benchmark model (column 2 of table 3) remain significant predictors. However, voluntary versus mandatory formats seem to be unrelated to effectiveness. Column 4 of table 4 shows the results on savings and retirement savings for low-income individuals reported in 31 studies. Signs and magnitude are similar to the benchmark estimation, but the only coefficients estimated with a small standard error are intensity per week and the teachable moment. Thus, qualitative results hold, but effect sizes on saving behavior for low-income individuals may be difficult to impact through the considered covariates.

Table 4. Explaining Heterogeneity in Effect Sizes for Subsamples by Type of Financial Behavior and Target Group

	(1) Borrow	(2) Borrow × low inc. clients	(3) Save	(4) Save × low inc. clients	(5) Budget	(6) Budget × low inc. clients
RCT	-0.136***	-0.100***	-0.002	-0.035		
	(0.022)	(0.026)	(0.045)	(0.058)		
TOT	0.089**	0.106**	0.090	0.074		
	(0.033)	(0.039)	(0.054)	(0.079)		
Delay	-0.000	-0.000*	0.000	-0.000	-0.001	-0.019
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.012)
1/SE	0.000	0.001**	-0.000	0.000	-0.003*	-0.007
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.005)
Intensity / week	0.003**	0.003**	0.003*	0.004**	0.037	0.595*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.031)	(0.308)
Duration	-0.000	-0.000	-0.001	0.000	-0.000	0.017
	(0.000)	(0.000)	(0.001)	(0.001)	(0.003)	(0.014)
Low income clients	-0.043**		-0.050**			
	(0.019)		(0.022)			
Years of schooling	-0.023***	-0.023***	-0.018***	-0.011	-0.020*	0.017
	(0.006)	(0.008)	(0.007)	(0.011)	(0.011)	(0.022)
Low/lower-mid. inc. econ.	-0.178***	-0.199***	-0.142***	-0.102		
	(0.052)	(0.067)	(0.045)	(0.066)		
Mandatory	-0.069**	-0.120***	-0.025	-0.010		
	(0.032)	(0.039)	(0.031)	(0.049)		
Teachable moment	0.100***	0.087***	0.084**	0.114*		
	(0.025)	(0.026)	(0.036)	(0.065)		
Constant	0.375***	0.326**	0.305***	0.147	0.361**	-0.685
	(0.087)	(0.114)	(0.091)	(0.165)	(0.134)	(0.524)
R^2	0.473	0.394	0.194	0.147	0.206	0.359
n (Studies)	32	20	67	31	20	11
n (Effect sizes)	100	73	166	91	40	27

Notes: Non-standardized coefficients from OLS regressions with clustered standard errors at the study-level in parentheses. We only include right hand side variables where differential information from at least two studies is available in the regressions. ***, ** and * denote significance at the one percent, five percent and ten percent level.
Source: Authors' analysis based on data sources discussed in the text.

Turning to the subsample of 20 studies on budgeting and record keeping behavior (column 5 of table 4), on which financial education yields the largest effects, we find that intensity is not significantly related to effect size. Additionally, all of the other signs and relative magnitudes of the coefficients remain similar to our benchmark estimation; however, with increased standard errors due to only 20 studies and 40 observations. Completing this exercise, we now examine determinants of treatment effects for the subsample of studies reporting on budgeting outcomes for low-income clients (column 6 of table 4). There are 11 studies in this subsample reporting 27 estimates. Again, qualitative results are similar and intensity now, again, is a marginally significant predictor of effect sizes on budgeting behavior.

Overall, we find that the positive effects from increased intensity appear to be driven by interventions focused on (retirement-) saving and borrowing behavior, whereas the timing and voluntary participation matter, especially for borrowing behavior. Thus, the financial behavior that is

hardest to impact (borrowing) needs special effort in the sense of increased intensity and timing the financial education intervention at a teachable moment.

VI. Robustness

The robustness tests cover eight different aspects and are reported in full in supplemental appendix S3. All of them confirm our qualitative findings. Here, we just mention these tests: (i) testing the average treatment effect with several alternative meta-regression models; (ii) repeating the parsimonious benchmark model without imputing missing values; (iii) running this model for studies about the US only; (iv) running this benchmark model with classroom studies only; (v) running this model with equal weight per study by either calculating one synthetic effect size per study or weighting effect sizes accordingly; (vi) running the benchmark specification with different empirical approaches; (vii) analyzing the influence of delay on effects; and (viii) testing a different definition of training intensity. Additionally, we further examine publication bias and possible heterogeneity in study quality in supplemental appendix S4 and use alternative econometric techniques that account for publication selection bias in supplemental appendix S3.

VII. Concluding Policy Discussion

This meta-analysis covers studies that potentially contribute to realizing policy objectives, such as improved financial literacy and changes in individual financial behavior. Due to this close link to economic policy, we discuss insights that have potential policy relevance in three steps:

General policy lessons: (i) the most important policy lesson from our research is that financial education can be effective. However, the field of financial education is not developed enough that established standards could be followed "blindly," rather the process of designing interventions needs careful attention due to large heterogeneity across program types and individual studies;(ii) interventions targeting improvements in financial literacy are quite successful as they achieve effectiveness similar to comparable education interventions in other domains. As financial literacy education basically aims at improving financial knowledge and awareness, it seems evidentiary that it works well in the classroom and at school (see e.g. Bruhn et al. 2016).

Improved financial literacy also has an indirect positive effect on financial behavior, although this indirect effect is small so that changes in financial behavior should also be addressed directly.

(iii) Education interventions targeting financial behavior have desired effects on average. Although these effects are economically rather small, they are statistically robust. Impacts on financial behavior

are higher if the intensity of education is increased and if financial education is offered at a teachable moment. The effects are smaller if "problematic" groups are addressed, such as low-income clients.

Policy lessons for subgroups. As the universe of studies covers widely diverse financial education interventions, we draw three lessons for more homogeneous groups: (i) regarding the country groups, education effects seem to be somewhat lower in *low and lower middle-income countries*. This is probably due to the disadvantageous institutional circumstances in these countries. A relative advantage in these countries, however, is that the general level of education (mean years of schooling in the population) is comparatively low so that marginal returns to additional domain-specific education are high. The lower opportunity costs of education may be a reason why mandatory participation conditions, such as school based programs, are less problematic and offering financial education at teachable moment appears to be of lesser importance in these countries;(ii) while problematic target groups, such as *low-income clients,* are more difficult to educate in general, the determinants of effective financial education are not different from the general population. If there is a difference, it appears that a teachable moment is relatively important, indicating that there is a particular need to get the attention of this target group; (iii) regarding the *outcomes* of financial education, improving debt related behavior is, on average, hardly successful. At the same time, mistakes can be rather consequential and the structure of many significant determinants is the same as for other financial behaviors, such that the general lessons may translate to this specific case; however, it needs much more input to reach economically significant results. Moreover, there is variation across studies revealing clear success cases, which suggests that it is useful to go down to the study level and learn from best practices. The effects on improving savings or budgeting behavior are much larger in magnitude than on borrowing.

Research on open policy issues. In order to improve financial education policies in the future we see three areas of urgent research: (i) we need quite generally *more reliable evidence on the effectiveness* of financial education interventions. Almost two-thirds of the evidence comes from the US, indicating that there are large gaps of evaluation elsewhere;(ii) regarding the documentation of impact evaluations within published reports, it would be very desirable to provide *more information about study and program characteristics* (see Miller et al. 2015). A straight-forward example is the quality of teacher training or implementation, which can make a crucial difference but is unknown in almost all studies (Brown et al. 2016). The same applies to the ways in which the curriculum is structured and implemented (see Drexler et al. 2014 as a notable exception); (iii) finally, in order to come closer to *welfare assessments,* information in two directions is needed: first, information about program costs is frequently missing. Thus, in terms of welfare, positive education effects could be balanced with the true costs of the intervention (see also Lusardi et al. 2016). Second, the discussion of effectiveness of financial education policy should also consider principal alternatives to financial

education in general. Such alternatives include limiting the kind of available products (choices), altering the choice architecture (e.g., Carroll et al. 2009), working with nudges (e.g., Thaler and Benartzi 2004; Willis 2011), considering the promotion of commitment devices (e.g., Brune et al. 2016), offering incentives (e.g., Saez 2009), or implementing more rigid consumer financial protection policies (cf. Campbell et al. 2011).

There are two arguments in favor of implementing financial education. First, the small average effect comes with low average intensity. More than 70 percent of our considered studies invest no more than two days in education, indicating that these measures may have only small effects, but also low costs. Second, the average small effect of financial education is accompanied by large heterogeneity, indicating that those offering financial education measures can still learn from best practice experiences, a development that is ongoing as evidenced by time trend of slowly increasing effectiveness documented in rigorous impact evaluation studies.

References

Agarwal, S., and B. Mazumder. 2013. "Cognitive Abilities and Household Financial Decision Making." *American Economic Journal: Applied Economics* 5 (1): 193–207. [CrossRef][10.1257/app.5.1.193]

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert, and C. Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236): 1260799.

Behrman, J.R., O.S. Mitchell, C.K. Soo, and D. Bravo. 2012. "How Financial Literacy Affects Household Wealth Accumulation." *American Economic Review: Papers and Proceedings* 102 (3): 300–4.

Berg, G., and B. Zia. 2013. "Harnessing Emotional Connections to Improve Financial Decisions. Evaluating the Impact of Financial Education in Mainstream Media." World Bank Policy Research Working Paper 6407.

Blue, L., P. Grootenboer, and M. Brimble. 2014. "Financial Literacy Education in the Curriculum: Making the Grade or Missing the Mark?" *International Review of Economics Education* 16 (Part A): 51–62.

Brown, M., J. Grigsby, W. van der Klaauw, J. Wen, and B. Zafar. 2016. "Financial Education and the Debt Behavior of the Young." *Review of Financial Studies* 29 (9): 2490–522.

Brune, L., X. Giné, J. Goldberg, and D. Yang. 2016. "Facilitating Savings for Agriculture: Field Experimental Evidence from Malawi." *Economic Development and Cultural Change* 64: 187–220.

Bruhn, M., L. de Souza Leao, A. Legovini, R. Marchetti, and B. Zia. 2016. "The Impact of High School Financial Education: Evidence from a Large-Scale Evaluation in Brazil." *American Economic Journal: Applied Economics* 8 (4): 256–95.

Bruhn, M., G.L. Ibarra, and D. McKenzie. 2014. "The Minimal Impact of a Large-Scale Financial Education Program in Mexico City." *Journal of Development Economics* 108: 184–9.

Bucher-Koenen, T., and M. Ziegelmeyer. 2014. "Once Burned, Twice Shy? Financial Literacy and Wealth Losses During the Financial Crisis." *Review of Finance* 18 (6): 2215–46.

Campbell, J.Y. 2006. "Household Finance." *Journal of Finance* 61 (4): 1553–604.

Campbell, J.Y., H.E. Jackson, B.C. Madrian, and P. Tufano. 2011. "Consumer Financial Protection." *Journal of Economic Perspectives* 25 (1): 91–114.

Card, D., J. Kluve, and A. Weber. 2010. "Active Labour Market Policy Evaluations: A Meta-Analysis." *Economic Journal* 120 (548): 452–77.

Card, D., J. Kluve, and A. Weber. 2015. "What Works? A Meta-Analysis of Recent Active Labor Market Program Evaluations." NBER Working Paper 21431.

Carpena, F., S. Cole, J. Shapiro, and B. Zia. 2015. "The ABCs of Financial Education. Experimental Evidence on Attitudes, Behavior, and Cognitive Biases." World Bank Policy Research Working Paper 7413.

Carroll, G.D., J.J. Choi, D. Laibson, B.C. Madrian, and A. Metrick. 2009. "Optimal Defaults and Active Decisions." *Quarterly Journal of Economics* 124 (4): 1639–74.

Choi, J.J., D. Laibson, and B.C. Madrian. 2010. "Why Does the Law of One Price Fail? An Experiment on Index Mutual Funds." *Review of Financial Studies* 23 (4): 1405–32.

———. 2011. "$100 Bills on the Sidewalk: Suboptimal Investment in 401(k) Plans." *Review of Economics and Statistics* 93 (3): 748–63.

Cohen, J. (1977). "Statistical power analysis for the behavioral sciences" (Rev. Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cole, S., T. Sampson, and B. Zia. 2011. "Prices or Knowledge? What Drives Demand for Financial Services in Emerging Markets?" *Journal of Finance* 66 (6): 1933–67.

Collins, J.M. 2013. "The Impacts of Mandatory Financial Education: Evidence from a Randomized Field Study." *Journal of Economic Behavior and Organization* 95: 146–58.

Collins, J.M., and C.M. O'Rourke. 2010. "Financial Education and Counseling—Still Holding Promise." *Journal of Consumer Affairs* 44 (3): 483–98.

Doi, Y., D. McKenzie, and B. Zia. 2014. "Who You Train Matters: Identifying Combined Effects of Financial Education on Migrant Households." *Journal of Development Economics* 109: 39–55.
Drexler, A., G. Fischer, and A. Schoar. 2014. "Keeping It Simple: Financial Literacy and Rules of Thumb." *American Economic Journal: Applied Economics* 6 (2): 1–31.
Duflo, E., and E. Saez. 2003. "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *Quarterly Journal of Economics* 118 (3): 815–42.
Fernandes, D., J.G. Lynch Jr, and R.G. Netemeyer. 2014. "Financial Literacy, Financial Education, and Downstream Financial Behaviors." *Management Science* 60 (8): 1861–83.
Freeman, S., S.L. Eddy, M. McDonough, M.K. Smith, N. Okoroafor, H. Jordt, and M.P. Wenderoth. 2014. "Active Learning Increases Student Performance in Science, Engineering, and Mathematics." *Proceedings of the National Academy of Sciences* 111 (23): 8410–5.
Fort, M., F. Manaresi, and S. Trucchi. 2016. "Adult Financial Literacy and Households' Financial Assets: The Role of Bank Information Policies." *Economic Policy* 31 (88): 743–82.
Fox, J., S. Bartholomae, and J. Lee. 2005. "Building the Case for Financial Education." *Journal of Consumer Affairs* 39 (1): 195–214.
Gathergood, J. 2012. "Self-Control, Financial Literacy and Consumer Over-Indebtedness." *Journal of Economic Psychology* 33 (3): 590–602.
Gerardi, K., L. Goette, and S. Meier. 2013. "Numerical Ability Predicts Mortgage Default." *Proceedings of the National Academy of Sciences* 110 (28): 11267–71.
Gertler, P.J., S. Martinez, P. Premand, L.B. Rawlings, and C.M. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington DC: World Bank Publications.
Gibson, J., D. McKenzie, and B. Zia. 2014. "The Impact of Financial Literacy Training for Migrants." *World Bank Economic Review* 28 (1): 130–61.
Grohmann, A., R. Kouwenberg, and L. Menkhoff. 2015. "Childhood Roots of Financial Literacy." *Journal of Economic Psychology* 51: 114–33.
Hastings, J.S., B.C. Madrian, and W.L. Skimmyhorn. 2013. "Financial Literacy, Financial Education, and Economic Outcomes." *Annual Review of Economics* 5: 347–73.
Imbens, G.W., and J.M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5–86.
Jappelli, T. 2010. "Economic Literacy: An International Comparison." *Economic Journal* 120 (548): F429–51.
Klapper, L., A. Lusardi, and P. van Oudheusden. 2015. "Financial Literacy Around the World: Insights from the Standard and Poor's Rating Services Global Financial Literacy Survey." http://gflec.org/initiatives/sp-global-finlit-survey/ ; last checked 07 August 2017.
Lipsey, M.W., and D.B. Wilson. 2001. *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.
Annamaria Lusardi, Pierre-Carl Michaud, and Olivia S. Mitchell, "Optimal Financial Knowledge and Wealth Inequality", *Journal of Political Economy* 125, no. 2 (April 2017): 431-477.
———. 2014. "The Economic Importance of Financial Literacy: Theory and Evidence." *Journal of Economic Literature* 52 (1): 5–44.
Lusardi, A., P.C. Michaud, and O.S. Mitchell. 2016. "Optimal Financial Knowledge and Wealth Inequality." *Journal of Political Economy*.
Miller, M., J. Reichelstein, C. Salas, and B. Zia. 2015. "Can You Help Someone Become Financially Capable? A Meta-Analysis of the Literature." *World Bank Research Observer* 30 (2): 220–46.
Muller, S.M. 2015. "Causal Interaction and External Validity: Obstacles to the Policy Relevance of Randomized Evaluations." *World Bank Economic Review* 29: S217–25.
Portnoy, D.B., L.A. Scott-Sheldon, B.T. Johnson, and M.P. Carey. 2008. "Computer-Delivered Interventions for Health Promotion and Behavioral Risk Reduction: A Meta-Analysis of 75 Randomized Controlled Trials, 1988–2007." *Preventive Medicine* 47 (1): 3–16.
Pritchett, L., and J. Sandefur. 2015. "Learning from Experiments When Context Matters." *American Economic Review: Papers and Proceedings* 105 (5): 471–5.

Saez, E. 2009. "Details Matter: The Impact of Presentation and Information on the Take-Up of Financial Incentives for Retirement Saving." *American Economic Journal: Economic Policy* 1 (1): 204–28.
Sayinzoga, A., E.H. Bulte, and R. Lensink. 2016. "Financial Literacy and Financial Behaviour: Experimental Evidence from Rural Rwanda." *Economic Journal* 126 (594): 1571–99.
Skimmyhorn, W. 2016. "Assessing Financial Education: Evidence from Boot Camp." *American Economic Journal: Economic Policy* 8 (2): 322–43.
Stango, V., and J. Zinman. 2009. "Exponential Growth Bias and Household Finance." *Journal of Finance* 64 (6): 2807–49.
Stanley, T.D. 2001. "Wheat from Chaff: Meta-Analysis As Quantitative Literature Review." *Journal of Economic Perspectives* 15 (3): 131–50.
Stanley, T.D., and H. Doucouliagos. 2012. *Meta-Regression Analysis in Economics and Business*. New York: Routledge.
Tennyson, S., and C. Nguyen. 2001. "State Curriculum Mandates and Student Knowledge of Personal Finance." *Journal of Consumer Affairs* 35 (2): 241–62.
Thaler, R.H., and S. Benartzi. 2004. "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy* 112: 164–87.
van Rooij, M., A. Lusardi, and R. Alessie. 2011. "Financial Literacy and Stock Market Participation." *Journal of Financial Economics* 101 (2): 449–72.
———. 2012. "Financial Literacy, Retirement Planning and Household Wealth." *Economic Journal* 122 (560): 449–78.
Vivalt, E. 2015. "Heterogeneous Treatment Effects in Impact Evaluation." *American Economic Review: Papers and Proceedings* 105 (5): 467–70.
von Gaudecker, H.-M. 2015. "How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice?" *Journal of Finance* 70 (2): 489–507.
Willis, L.E. 2011. "The Financial Education Fallacy." *American Economic Review: Papers and Proceedings* 101 (3): 429–34.
Xu, L., and B. Zia. 2012. "Financial Literacy Around the World: An Overview of the Evidence with Practical Suggestions for the Way Forward." World Bank Policy Research Working Paper 6107.
Zinman, J. 2015. "Household Debt: Facts, Puzzles, Theories, and Policies." *Annual Review of Economics* 7: 251–76.

Appendix

to accompany

"Does financial education impact financial literacy and financial behavior, and if so, when?"

Appendix S1: Supplementary material

Appendix S2: Comparison of our dataset and results to previous meta-analyses

Appendix S3: Robustness checks

Appendix S4: Publication bias and heterogeneity of study quality

Appendix S5: Overview of studies included in the statistical meta-analysis

Appendix S6: References for studies included in the statistical meta-analysis

Appendix S1: Supplementary material

This Appendix S1 contains two kinds of information: First, there are three tables (Table S1.1 to Table S1.3) and two figures (Figure S1.1 and Figure S1.2), which are referred to in the main text, mainly in the earlier sections.

Second, there is a longer documentation about "Additional information on selection of studies and extraction of effect size estimates and study descriptors." This documentation provides deeper information that complements Section 3.1 (Selection of studies) and Section 3.2 (Extraction of effect size estimates and study descriptors) of the main text.

Table S1.1: Summary of financial education studies by publication date and country

	Number of studies (1)	Percent of sample (2)	
A By publication date			
1999	2	1.59	
2000	0	0.00	
2001	5	3.97	
2002	1	0.79	
2003	4	3.17	
2004	3	2.38	
2005	6	4.76	
2006	5	3.97	
2007	6	4.76	
2008	6	4.76	
2009	8	6.35	
2010	10	7.94	
2011	7	5.56	
2012	15	11.9	
2013	9	7.14	
2014	11	8.73	
2015	15	11.9	
2016	13	10.32	
B By country of intervention			*Income*
Australia	2	1.59	High
Bosnia and Herzegovina	1	0.79	Upper-middle
Brazil	1	0.79	Upper-middle
China	1	0.79	Upper-middle
Dominican Republic	1	0.79	Upper-middle
Germany	1	0.79	High
Ghana	1	0.79	Lower-middle
Hong Kong, China	1	0.79	High
India	8	6.35	Lower-middle
Indonesia	2	1.59	Lower-middle
Italy	7	5.56	High
Kenya	1	0.79	Lower-middle
Malawi	1	0.79	Low
Mexico	1	0.79	Upper-middle
Mozambique	1	0.79	Low
New Zealand	2	1.59	High
Pakistan	1	0.79	Lower-middle
Qatar	1	0.79	High
Rwanda	1	0.79	Low
Singapore	1	0.79	High
South Africa	1	0.79	Upper-middle
Spain	1	0.79	High
Sri Lanka	1	0.79	Lower-middle
Tanzania	2	1.59	Low
USA	83	65.87	High
Uganda	2	1.59	Low
Low inc. econ.	7	5.5	
Lower-middle inc. econ.	14	11.11	
Upper-middle inc. econ.	6	4.76	
High inc. econ.	99	78.57	
Total	126	100	

Notes: Country group classifications refer to 2015 World Bank data on GNI per capita (Atlas method).

Table S1.2: Overview of coded outcomes and definitions

	Outcome category	Definition	Freq.
Financial literacy (190 estimates)			
A	*Financial knowledge (+)*	Raw score on financial knowledge test Indicator of scoring above a defined threshold Indicator of solving an item correctly	190 (100%)
Financial behaviors (349 estimates)			
B	*Borrowing & debt management behavior*		100 (28.65%)
1)	Reduction of loan default within a certain time-frame (+)	Binary indicator	
2)	Reduction of delinquencies within certain time frame (+)	Binary indicator	
3)	Better credit score (+)	Continuous measure of credit score	
4)	Reduction in informal borrowings (+)	Binary indicator of informal loan or reduction in number of informal loans	
5)	Lower cost of credit / interest rate (+)	Sum of real interest amount or interest rate and (if applicable) cost of fees	
6)	Any debt (-) / (+) (depending on intervention goal)	Binary indicator Binary indicator	
7)	Any formal loan (+)	Continuous measure of borrowed amount	
8)	Total amount borrowed (-) / (+) (depending on intervention goal)		
9)	Total outstanding debt (-) / (+) (depending on intervention goal)	Continuous measure of total debt	
10)	Better borrowing index (+)	Study-specific index of survey items to measure borrowing amount, frequency, and repayment	
11)	Uses credit card up to limit (-)	Binary indicator	
C	*Budgeting & planning behavior*		40 (11.46%)
1)	Having a written budget (+)	Binary indicator	
2)	Positive sentiment toward budgeting (+)	Binary indicator	
3)	Having a financial plan (+)	Binary indicator	
4)	Keeping separate records for business and household (+)	Binary indicator	
5)	Seeking information before making financial decisions (+)	Binary indicator	
6)	Self-rating of adherence to budget (+)	Study-specific scale	
D	*Saving & retirement saving behavior*		166 (47.56%)
1)	Total savings held (+)	Continuous measure of savings amount or categorical variable indicating amount within range	
2)	Savings rate or savings within timeframe (+)	Savings relative to income Amount over defined time-frame	
3)	Savings index (+)	Study-specific index of survey items designed to measure savings amount and frequency	*-continued-*
4)	Any savings (+)	Binary indicator	

	5) Has formal bank (savings) account (+)	Binary indicator	
	6) Investments into own or other business (stocks) (+)	Continuous measure of amount invested	
	7) Holds any stocks or bonds (+)	Binary indicator	
	8) Has any retirement savings (+)	Binary indicator	
	9) Participates in retirement savings plan (e.g. 401k) (+)	Binary indicator	
	10) Amount of retirement savings (+)	Continuous measure of retirement savings amount	
	11) Retirement savings rate (+)	Retirement savings relative to income	
	12) Positive sentiment towards investing funds (+)	Binary indicator	
	13) Reduction of excess risk in retirement fund (+)	Continuous measure of retirement savings amount allocated to risky assets	
	14) Reduction of cost of savings product (fees paid) (+)	Continuous measure of fee amount paid	
	15) Increase in contribution rate to retirement savings plan (+)	Indicator of increase or continuous measure of amount increase	
	16) Net wealth (+)	Continuous measure of net wealth	
E	*Insurance & risk mitigation behavior*		16 (4.59%)
	1) Any formal insurance (+)	Binary indicator	
	2) Having a diversified portfolio (+)	Numbers of assets in portfolio; Standard deviation of returns in portfolio	
F	*Remittance behavior*		16 (4.59%)
	1) Lower cost of remittance product (+)	Continuous measure of cost or binary choice of lower cost product	
	2) Lower remittance frequency and higher amount (lower cost) (+)	Measure of remittance frequency within timeframe and continuous amount remitted	
	3) More control over remitted funds (+)	Study-specific scale to measure control over remitted amount	
G	*Bank account behavior*		11 (3.15%)
	1) Has formal bank (checking) account (+)	Binary indicator	
	2) Opens formal account within certain time frame	Binary indicator	
	3) Uses formal bank account	Binary indicator	

Notes: When necessary, outcomes are reverse-coded so that positive signs reflect positive financial education treatment effects (i.e. when the dependent variable is coded as the probability of default, we transform this to the reduction in probability of default in order to be able to assign a positive sign).

Table S1.3: Summary of estimated financial education impacts

Outcome	Significance at 5%			Significance at 10%			Average effect size (SE)
	Negative	*Insig.*	*Positive*	*Negative*	*Insig.*	*Positive*	
A Effects on financial literacy							
Fin. literacy	1 (0.53%)	72 (37.89%)	117 (61.58%)	2 (1.05%)	62 (32.63%)	126 (66.32)	0.263*** (0.414)
B Effects on financial behavior							
Fin. behavior	8 (2.29%)	215 (61.60%)	126 (36.10%)	18 (5.16%)	181 (51.86%)	150 (42.98%)	0.086*** (0.012)
Borrowing	5 (5.00%)	80 (80.00%)	15 (15.00%)	10 (10.00%)	70 (70.00%)	20 (20.00%)	0.023 (0.014)
Budgeting & planning	0 (0.00%)	15 (37.5%)	25 (62.50%)	1 (2.50%)	10 (25.00%)	29 (72.50%)	0.207*** (0.053)
Saving	2 (1.67%)	61 (50.83%)	57 (47.50%)	6 (5.00%)	49 (40.83%)	65 (54.17%)	0.108*** (0.017)
Retirement Saving	0 (0.00%)	22 (47.83%)	24 (52.17%)	0 (0.00%)	17 (36.96%)	29 (63.04%)	0.108*** (0.034)
Insurance	0 (0.00%)	13 (81.25%)	3 (18.75%)	0 (0.00%)	12 (75.00%)	4 (25.00%)	0.045 (0.024)
Bank account behavior	0 (0.00%)	10 (90.91%)	1 (9.09%)	0 (0.00%)	10 (90.91%)	1 (9.09%)	0.003 (0.027)
Remittance behavior	1 (6.25%)	14 (87.50%)	1 (6.25%)	1 (6.25%)	13 (81.25%)	2 (12.50%)	0.035 (0.046)

Notes: Average effect sizes are estimated via OLS with standard errors clustered at the study-level in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level.

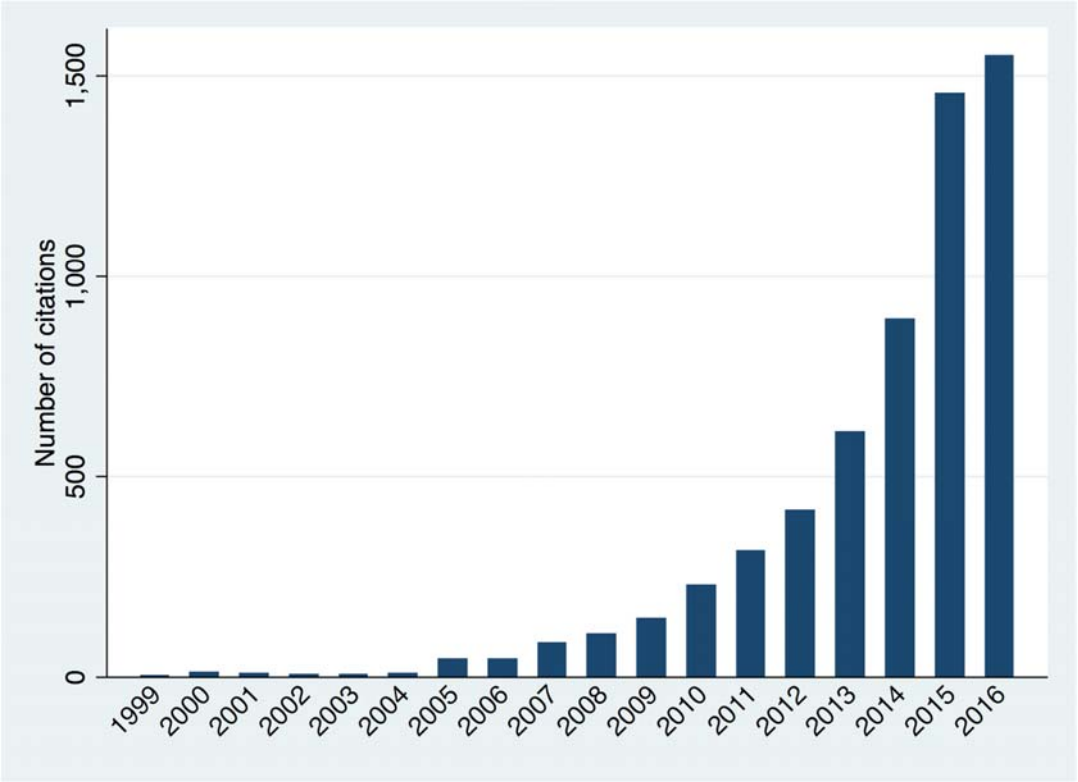


Figure S1.1: Citations of published items with the keyword financial literacy per year, source: SSCI

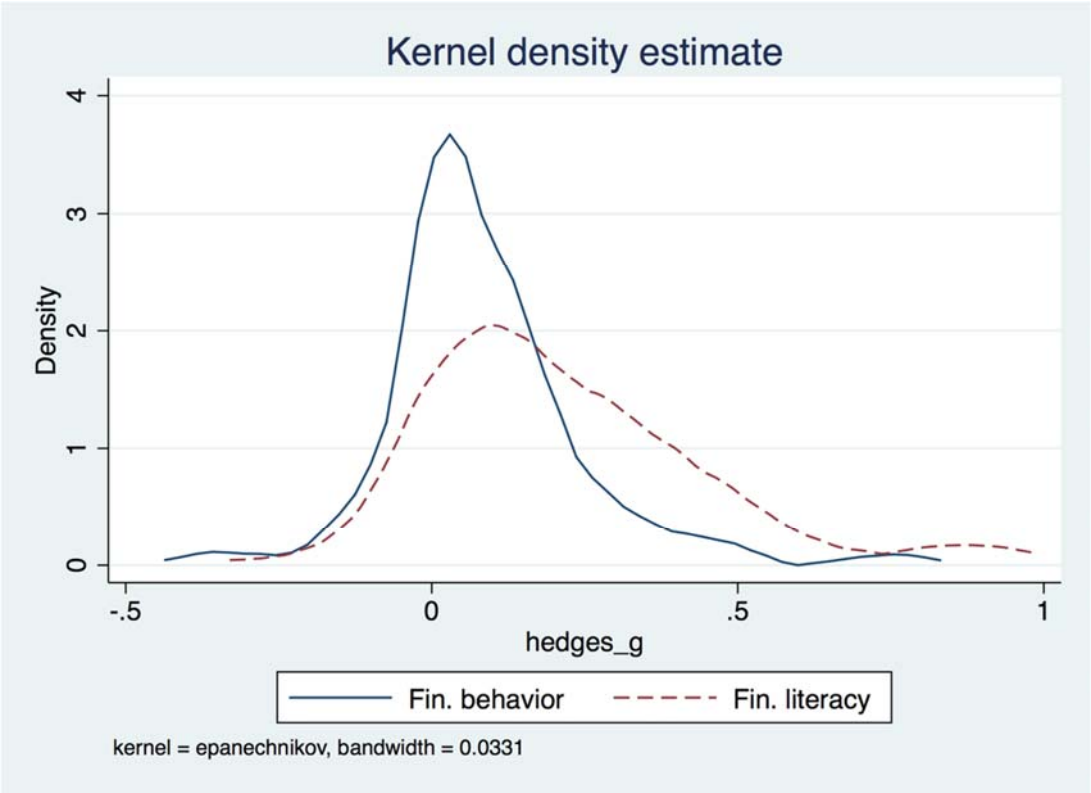


Figure S1.2: Kernel-density estimates of effect sizes by outcome (for Hedge's g<1)

Additional information on selection of studies and extraction of effect sizes estimates and study descriptors.

Selection of studies. We follow the established meta-analytical protocol (cf. Lipsey and Wilson 2001, p.23; Stanley 2001, p.143; Stanley and Doucouliagos 2012; Stanley et al. 2013). This starts with systematically searching the relevant databases for the most common keywords in order to aggregate a large sample of potentially eligible studies to be included in our meta-analysis. Keywords are (i) financial literacy; (ii) financial knowledge; (iii) financial education; (iv) financial capability; and (v) combinations of these keywords with "intervention." To minimize publication bias and capture the broadest sample of studies possible, we systematically search not only the relevant databases for published records (e.g. ISI, Business Source Premier via EBSCO Host, JStor) but also for registered trials, working papers, and informal research reports (e.g. AEA RCT-registry, SSRN, Fin. Lit. E-Journal, RePEC, NBER, Worldbank eLibrary). All records from recent systematic accounts of the literature (Fernandes et al. 2014; Miller et al. 2015) are included in our initial pool of studies. In addition, we screen the references of narrative literature reviews (Fox et al. 2005; Collins and O'Rourke 2010; Willis 2011; Xu and Zia 2012; Hastings et al. 2013; Blue et al. 2014; Lusardi and Mitchell 2014).

This search resulted in over 500 potentially relevant published journal-articles and over 600 results from working paper databases with some apparent overlap. We stopped collecting articles from these databases in October 2016.

From this collection, we drop studies that do not meet our three criteria of inclusion: (i) Reporting on impacts of an exogenous educational intervention designed to strengthen the participants' financial literacy and/or leading to behavioral change in the area of personal finance; (ii) providing a quantitative assessment of intervention impact that allows coding an effect size statistic (g) and its standard error; and (iii) relying on an observed counterfactual in

the estimation of intervention impacts. Consequently, we only include experimental studies with sufficient information on intervention outcomes in our analysis, i.e. RCTs, quasi-experiments, and natural experiments (see below for coding of studies). Where necessary information was partially missing, we consulted additional online resources related to the article or contacted the authors of the primary studies via e-mail.

This selection-process results in a final sample of 126 independent intervention studies that report 539 effect sizes. Of these, 90 studies report 349 effect sizes on financial behavior, and 67 studies report 190 effect sizes on financial literacy. Among these 90 plus 67 studies, there are 31 studies reporting effect sizes for both financial literacy and behavior. Our selection of studies covers 126 independent interventions from 1999 through 2016. Table S1.1 shows the composition of our sample of studies by the date of publication (*Panel A*) and the country in which the intervention took place (*Panel B*). While most interventions took place in the U.S. and other OECD countries, 21.4% of studies were conducted in low- or middle-income countries. The sample is comprised of 51 RCTs and 75 quasi-experiments. RCTs are rare in the early years of the literature, but the share has risen dramatically, with the majority of studies conducted from 2011 onward being randomized evaluations (see Figure 1 in the main text).

Extraction of estimates. The next step in our meta-analytic process is to extract effect size estimates from the statistical data reported in the primary studies. Our analysis aggregates treatment effects of financial education interventions on two main categories of outcomes. First, we code the effect of financial education on *financial literacy* (i.e. a measure of performance on a financial knowledge test) since knowledge development is the primary goal of financial education (Hastings et al. 2013; Lusardi and Mitchell 2014). We do not include self-assessments of changes in financial knowledge as an outcome.

Second, we code treatment effects of financial education on *financial behaviors.* These behaviors can be further disaggregated into the following categories: Borrowing, savings and

retirement saving, budgeting and planning, insurance, as well as remittances. Table S1.2 provides an overview of the categories and definitions of effect size estimates by outcome type.

We code all available effect sizes per study on cognitive (financial knowledge) and behavioral outcomes. We include multiple estimates per study if multiple outcomes, time-points, or treatments are reported. We only extract main (average treatment) effects reported in the papers. Thus, we do not code estimates reported in the "heterogeneity-of-treatment-effects-section" within papers, such as sample splits or interaction-effects of binary indicators (e.g. gender, income, ability, …) with the treatment indicators. If results are only reported in a disaggregated manner (only effects on subsamples), we perform a within study (random-effects) meta-analysis (DerSimonian and Laird 1986) to generate an inverse-variance-weighted average effect size to proxy the main effect. Additionally, we aim to capture only non-redundant effect sizes per paper (i.e. we do not include effect sizes for the same intervention on the same outcome reported in the robustness-section). The number of coded estimates per study ranges from 1 to 87 estimates. We show in the Appendix S3 (robustness checks) that giving each study equal weight by creating a single synthetic effect size per study through a within-study meta-analysis or alternatively by weighting each observation by the inverse number of effect-size estimates contributed by each study yields similar results.

In addition to the coding of all possible estimates of effect sizes (g) and their standard errors of financial education treatment on financial literacy or financial behavior (cf. Section 2), we develop a coding protocol to extract potentially relevant information about the study (study descriptors) that may serve as predictor variables explaining the variability in effect sizes. Specifically, we aim at extracting data on (i) research design and measurement of dependent variables; (ii) the intensity of education; (iii) the sample/target group of the intervention; and (iv) the details of the intervention itself, such as channel, setting, and participation conditions. Coding of the included study reports was completed by the authors of

this paper and two research assistants who were trained using the guidelines by Lipsey and Wilson (2001, p.88). Overall intercoder reliability is high and data collection for most of the variables concerning the setting, participants, and research design of the primary studies was straightforward. However, key details of the underlying educational intervention are often missing or underreported in the research reports. If information is only partially missing authors were asked to provide these details via e-mail.

References in Appendix S1

Blue, L., Grootenboer, P., and Brimble, M. (2014). Financial literacy education in the curriculum: Making the grade or missing the mark? *International Review of Economics Education*, 16, Part A(0): 51–62.

Collins, J.M. and O'Rourke, C.M. (2010). Financial education and counseling—still holding promise. *Journal of Consumer Affairs*, 44(3): 483–498.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3): 177–188.

Fox, J., Bartholomae, S., and Lee, J. (2005). Building the case for financial education. *Journal of Consumer Affairs*, 39(1): 195–214.

Fernandes, D., Lynch, Jr., J.G., and Netemeyer, R.G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, 60(8): 1861–1883.

Hastings, J.S., Madrian, B.C., and Skimmyhorn, W.L. (2013). Financial literacy, financial education, and economic outcomes. *Annual Review of Economics*, 5: 347–373.

Lusardi, A. and Mitchell, O.S. (2014). The economic importance of financial literacy: theory and evidence. *Journal of Economic Literature*, 52(1): 5–44. Lipsey, M.W. and Wilson, D.B. (2001). *Practical meta-analysis.* Sage, Thousand Oaks, CA.

Miller, M., Reichelstein, J., Salas, C., and Zia, B. (2015). Can you help someone become financially capable? A meta-analysis of the literature. *World Bank Research Observer*, 30(2): 220–246.

Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of Economic Perspectives*, 15(3): 131–150.

Stanley, T. D. and Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*, Routledge, New York, NY.

Stanley, T., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., Nelson, J. P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R. S., and Rost, K. (2013). Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys*, 27(2):390–394.

Willis, L.E. (2011). The financial education fallacy. *American Economic Review: Papers and Proceedings*, 101(3): 429–434.

Xu, L., and Zia, B. (2012). Financial literacy around the world: an overview of the evidence with practical suggestions for the way forward. *World Bank Policy Research Working Paper 6107.*

Appendix S2: Comparison of our dataset and results to previous meta-analyses

There are two earlier meta-analyses about financial education: The study by Miller et al. (2015) covers 19 papers due to its extremely restrictive selection criteria. Thus, most similar to our work is the study by Fernandes et al. (2014), which covers 90 effect sizes from financial education reported in 77 papers. Despite an overlap of 44% with their sample of studies, our research differs in four ways which explains our new results: (i) most important is that we analyze determinants of program effectiveness in a broader way by applying respective coding. (ii) Then we code the various outcomes per study and their respective effectiveness. Moreover, (iii) we cover recent and mostly randomized experiments providing evidence of effective interventions; and (iv) we cover additional studies focusing exclusively on financial literacy as the outcome variable. We aim to elaborate on these comparisons in this part of the Appendix.

Comparison of studied samples. Our selection-process (see Appendix S1) led us to a final sample of 126 independent intervention studies that report 539 effect sizes. Of these, 90 studies report 349 effect sizes on financial behavior, and 67 studies report 190 effect sizes on financial literacy. Among these 90 plus 67 studies, there are 31 studies reporting effect sizes on both financial literacy and behavior. The sample is comprised of 51 RCTs and 75 quasi-experiments.

As mentioned, Miller et al. (2015) select 19 intervention-studies for their statistical meta-analysis. Their main inclusion criterion is that interventions report on identical outcomes. This limits their analysis to sample sizes of four to six studies (and estimates) per outcome. While informative of magnitude and significance of effect sizes on identical outcomes, such an approach prevents a detailed investigation into the sources of heterogeneity, given the very limited number of studies available. However, we note that the results for size, direction, and significance of the main behaviors studied in Miller et al. 2015 are in line with our results (see Figure 2 in the main text).

Fernandes et al. (2014), with 77 papers selected, cover 90 effect sizes (15 RCTs and 75 quasi-experiments) of "manipulated financial literacy" (cf. Fernandes et al. 2014, p.1863). Of their 77 papers, 55 are also part of our sample. We exclude 22 single-group pre-posttest and quasi-experimental papers because they either do not analyze education interventions (but other personal finance related programs, e.g. match incentives), report only aggregate measures of self-reported financial behavior, wellbeing or self-efficacy, or because it is not feasible to calculate a meaningful effect size statistic. In addition, we include 35 recent studies that were not previously available. Moreover, we consider another 36 studies examining the impact of financial education on financial literacy but neglecting possible impacts on financial behavior. These differences explain the mentioned overlap of 44% regarding studies.

Comparison of estimation results. We estimate the average treatment effect of educational interventions on *financial behaviors* to be statistically highly significant (g=0.086, p=0.000, n=349). Although the average treatment effect of 0.086 is small in magnitude, there exists a measurable and robust impact of financial education on various kinds of financial behavior. In comparison, Fernandes et al. (2014) estimate the summary effect of financial education on financial behavior to be roughly g=0.066. However, the authors use averaged effect sizes per paper and weight each observation with its average inverse variance. In order to obtain a better comparison with that study, we exactly apply their method (random effects meta-regression) with synthetic effect sizes per study to our sample of studies. This provides an average (weighted) effect size of g=0.079 (p=0.000, n=90) (see Table S3.1 in Appendix S3). Thus, our estimate of a summary effect for the literature is not too different from theirs.

To investigate the potential source of this difference, we estimate the weighted average effect size among those recent studies that are not included in Fernandes et al. (2014). Indeed, we find that there is a larger average effect of financial education on financial behavior in this sample (g=0.13). This indicates that the new studies covered in our meta-analysis are the main

source of difference. Diving deeper into this issue, we find that Fernandes et al. (2014) estimate extremely small average effect sizes for their sample of 15 RCTs. Our broader sample of randomized experiments, however, leads to a much more positive assessment. In line with this observation, the effect size of financial education on financial behavior documented in RCTs seems to increase over time, indicating a positive time trend in effect sizes: a regression of effect size on year of study publication results in a statistically highly significant coefficient (b=0.014, SE=0.004). This moderate, positive time-trend is an important element in explaining our positive result about the effect of financial education on financial behavior.

Turning to the result concerning the treatment effect of financial education on *financial literacy* (measured through knowledge assessments), we estimate the average impact of financial education on financial literacy to be g=0.263 (p=0.000, n=190). Thus, our analysis of a comprehensive sample of studies (n=67) leads to a positive assessment of the effectiveness of financial education on financial literacy. This education explains 1.7% of the variance in financial knowledge and, thus, appears only slightly less effective than educational interventions in other domains, such as math and science instruction (cf. Fernandes et al. 2014, p.1867). Our positive result is in remarkable contrast to Fernandes et al. (2014, p.1867), who find that financial education only explains 0.4% of the variance in financial literacy and state accordingly that, "financial education yields surprisingly weak changes in financial knowledge presumed to cause financial behavior." However, this result seems a bit fragile as it is based on only 12 studies and cannot, obviously, be replicated in our larger sample of studies (cf. Fernandes et al. 2014, p. 1867).

References in Appendix S2

Fernandes, D., Lynch, Jr., J.G., and Netemeyer, R.G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, 60(8): 1861–1883.

Miller, M., Reichelstein, J., Salas, C., and Zia, B. (2015). Can you help someone become financially capable? A meta-analysis of the literature. *World Bank Research Observer*, 30(2): 220–246.

Appendix S3: Robustness checks

Appendix S3 contains eight kinds of robustness checks: (i) we estimate the (weighted) average treatment effect of financial education on financial behavior using five alternative meta-regression models for continuous effect sizes; (ii) we show results without imputing missing values; (iii) we run our benchmark analysis with the subsample of studies conducted in the USA only; (iv) we run our benchmark analysis with the subsample of classroom financial education studies only; (v) we give each study the same weight in the analysis by creating one synthetic effect size per study or, alternatively, assigning a weight of the inverse number of observations contributed by each study to each estimate within a given study; (vi) we re-estimate our multivariate analysis using eleven alternative meta-regression models; (vii) we look for heterogeneous impacts depending on the delay in measurement of outcomes; and, lastly, (viii) we test a different operationalization of training intensity.

(i) Summary of treatment effects on financial behavior under various models. Table S3.1 shows the estimated (weighted) average effect size of financial education treatment on financial behavior outcomes for six alternative models. We first perform an analysis on the full sample (Panel A) and disaggregate our sample further into RCTs only (Panel B) and a subsample containing only quasi and natural experiments (Panel C).

<Table S3.1 about here>

Column (1) repeats the OLS results, while Column (2) shows results with a single synthetic (weighted average) effect size per study. Column (3) shows results for random effects meta-regression (DerSimonian and Laird 1986) with inverse variance weights, synthetic effect sizes per study, and Knapp and Hartung (2003) adjusted standard errors. This is common in meta-analyses in other disciplines (such as clinical trials) and thus serves as a further check of the sensitivity of our results to the estimation strategy. This approach assigns weights for each

study based on the inverse variance of the within study measurement error plus the between study variance (tau squared) ($w_i = \frac{1}{V_{Yi}+\tau^2}$). Thus we define our meta-analytic model as

$$g_i = x_i\beta + u_i + \epsilon_i \tag{6}$$

where

$$u_i \sim N(0, \tau^2) \tag{7}$$

and

$$\epsilon_i \sim N(0, \sigma_i^2) \tag{8}$$

Here g_i is defined as the effect size estimate of study *i*, $\sigma_i$is the corresponding standard error, $\tau^2$is the between study variance in true effects, and $x_i\beta$ is a vector of study level covariates (including an intercept). We estimate this model using either method of moments (DerSimonian and Laird 1986) or alternatively restricted maximum likelihood or empirical bayes.

Column (4) reports estimations based on a GLS random-effects model. If one assumes that the between-study heterogeneity cannot readily be explained by the observable characteristics included, $x_{ij}\beta$ (i.e. due to unobserved heterogeneity in implementation quality), one has to incorporate unobservable characteristics through random effects into the model (cf. Cho and Honorati 2014). Thus, including an effect capturing unobservable characteristics of the study, the meta-analytic model is defined as:

$$g_{ij} = x_{ij}\beta + \theta_{ij} + \epsilon_{ij} \tag{5}$$

where g_{ij} is the impact (continuous effect size) of a financial education intervention on outcome *i* reported in study *j*, $x_{ij}\beta$ is a vector of observable covariates, θ_{ij} is a random effect of unobservable study characteristics and $\epsilon_{ij}$is an error term independent of $x_{ij}\beta$ and θ_{ij}.

Column (5) shows results for full pooling unrestricted weighted least squares using the inverse standard error (precision) as weights (cf. Stanley and Doucouliagos 2012, 2015).

Finally, Column (6) shows results from robust variance meta-regression with dependent effect sizes (see Tanner-Smith and Tipton 2014).

Reassuringly the direction is positive and statistical significance is found for all of the considered models and sample splits. Additionally, the magnitude of the coefficient is similar; however differences in detail do exist: The most common meta-analysis model is presented in Column (3), which is also the model that Fernandes et al. (2014) and Miller et al. (2015) use for their analyses. These models compare favorably to our main results discussed in the paper relying on unrestricted ordinary least squares using multiple estimates per study and clustering the standard errors at the study-level. In contrast, unweighted random effects GLS leads to a higher estimate of the average treatment effect (Column 4). This approach is used previously by Cho and Honorati (2014). The smallest estimate is reported in Column (5): By relying on unrestricted weighted least squares, very large studies with extremely small standard errors, which are most often quasi-experimental, receive extreme weight in the calculation of the summary effect. From our point of view, it does not seem ideal to discount comparatively smaller studies (which often still have sample sizes of over 1000 individuals) with high internal validity (RCTs) as strongly as this approach does. Thus, if one incorporates weights based on the standard error or variance of estimates, it seems advisable to account for between study heterogeneity through random effects as discussed above and presented in columns (3) and (6). Finally, column (6) presents results applying a recently developed method that accounts for dependency among effect sizes (multiple, correlated estimates per study) (see Hedges et al. 2010; Tanner-Smith and Tipton 2014). Again, results are in line with our main results, although with deflated expectations about the average effect in the whole sample of studies. This estimate is also in line with the magnitude of the result presented in Fernandes et al. (2014), however, our assessment about the effectiveness of 40 RCTs on financial behavior remains to be strikingly different to the evidence synthesized by Fernandes et al. (2014).

(ii) Conservative handling of missing data. Next, we turn to estimations of complete cases only, in order to test the robustness of our results using imputed default categories or mean values for missing observations. Column (1) in Table S3.2 reports OLS meta-regression results for complete cases only. These results correspond to the results presented in Table 3 of the manuscript but show larger standard errors for some of the variables, however, turning none of the main explanatory variables insignificant. This result strongly supports the conclusions drawn from estimations with a large number of studies in the sample.

<Table S3.2 about here>

(iii) US only subsample. Then we consider only studies conducted in the U.S., since these account for 65.87% of the studies and 42.67% of the effect size estimates in our sample (column 2 of Table S3.2). Again, our results are near identical to the estimation in Table 3. However, the standard error for the covariate for low-income clients increases and turns this result insignificant while maintaining its magnitude and sign.

(iv) Classroom trainings only. Further, we consider only studies reporting on classroom trainings as interventions (column 3). Again, our results are near identical to the estimation in Table 3. However, the standard error for the covariate for mandatory courses increases.

(v) Equal study weights. Much of the meta-analysis literature in other fields than economics uses effect size models where each study contributes only one synthetic effect size to the meta-regression analysis. This procedure assures that the assumption of independent estimates is not violated. There are different options to provide such a single effect. Some suggest only using the most robust results in a primary study (cf. Cho and Honorati 2014, p. 119). The textbook literature on meta-analysis, however, tends to recommend creating a synthetic effect size per study by using the average (or weighted average) effect across multiple outcomes (cf. Lipsey and Wilson 2001).

We follow this approach here for the purpose of robustness exercises, but we point at the major disadvantage that effects heading in opposite directions within one study may be cancelling each other out. Column (4) of Table S3.2 shows results for such an approach. The signs and magnitudes of our coefficients are very similar to the model with multiple non-synthetic effect sizes per study and standard errors clustered at the study-level. However, in the estimation based on this sample, the standard errors increase, thus leading to insignificant covariates in three cases: RCT, intensity per week, and low-income clients. Since this approach works with much less information than would be otherwise available, we conclude that qualitatively this check also confirms our main findings derived from the larger sample of available effect sizes.

Finally, in column (5) we give each study equal weight by assigning the inverse number of estimates per study as weights for each effect size observation within a study. This yields very similar results to the approach in column (4).

(vi) Alternative meta-regression models. Here we discuss the use of alternative statistical regression models in the estimation of predictors of intervention impact.

(Ordered) probit models for sign and significance. In column (1) of Table S3.3 we apply a probit-regression on an indicator variable of statistically significant effect estimates (at the 5%-level). This is a departure from earlier analyses because we now neglect the size of effects and only consider their statistical significance. Following the approach applied by Card et al. (2010, 2015) and Cho and Honorati (2014), we code the sign and significance for each impact estimate reported in the primary studies. This indicator of intervention success has the advantage that it is easily interpretable and neutral to the unit of the outcome variable. However, it only captures the direction and significance of an effect, unlike the standardized mean difference which preserves its magnitude (cf. Stanley and Doucouliagos 2012, p. 6). Using this approach, we construct a binary dependent variable taking the value 0 if the primary study

impact estimate t-statistic is smaller than 1.96 and taking the value 1 if $t \geq 1.96$. Additionally, we extend this approach and construct an ordered categorical variable that can take three values of -1 if $t \leq -1.64$, 0 if $t \geq -1.64$ and $t \leq 1.64$, and 1 if $t \geq 1.64$. Thus, we distinguish between significant negative, insignificant, and significant positive estimates at the 10%-level because there are hardly negative estimates at the 5%-level (see Table S1.3 in the Appendix S1).

<Table S3.3 about here>

We observe that mostly the sign and significance of the logged odds correspond with the model using a continuous measure of effect size reported in Table 3, column (2). However, estimated standard errors differ, as the coefficients for TOT, intensity, and mandatory are now insignificant – probably resulting from reduced variance in the dependent variable in comparison to the use of continuous effect sizes.

In column (2) we extend this approach and estimate an ordered probit model where the dependent variable consists of three ordered categories that distinguish between significant negative, insignificant and significant positive estimates at the 10%-level of financial education impact. This leads to a very similar assessment of predictor sign and magnitude as in our benchmark model in Table 3, column (2), but again slightly different estimates for the standard errors, with intensity, however, being a significant predictor in this estimation again.

GLS random effects regression. Next, we check whether controlling for unobservables affects our results. The results in column (3) show coefficients from a GLS random effects regression based on the assumptions discussed in equation 5. This estimation almost entirely matches the results of the benchmark model shown in Table 3, column (2) with the exception of an increased standard error for mandatory financial education.

Unrestricted weighted least squares. Next, we turn to an alternative unrestricted weighted least squares approach. In column (4) we weight each effect size with its inverse standard error (1/SE) and account for publication selection bias by including the standard error (SE) of each

estimate as a covariate (as suggested by Stanley and Doucouliagos 2012). The results show that our results, again, largely match the results of the ordinary least squares estimations, however, the predictor for mandatory courses is now insignificant. In column (5) we redo this analysis and use the inverse variance as weights and include the variance as a covariate in the analysis to account for publication selection bias. This estimation, while qualitatively similar, shows no negative effects (due to increased estimated standard errors) for low-income clients, and mandatory courses.

Random effects meta-regression (DerSimonian and Laird 1986). Table S3.4 shows our preferred specification for three different estimators of random-effects meta-regression models with and without Knapp and Hartung (2003) corrected standard errors, respectively. Using method of moments (columns 1 and 2), we find that our results are similar to our benchmark model using OLS in Table 3, column (2), with the exception of increased standard errors, especially when applying the correction suggested by Knapp and Hartung (2003), for the coefficients for low-income economies, low-income individuals and intensity per week, which are now statistically insignificant. Turning to the alternative estimators (restricted maximum likelihood, and empirical bayes), we find that these results are again nearly identical. Overall, we conclude that the pattern in sign and magnitude (including most standard errors) of our main explanatory variables are confirmed under various random effect meta-regression models, however with a more positive assessment of the intervention impact in low and lower-middle income economies and for low-income individuals, as well as a positive but insignificant estimate of intensity per week.

<Table S3.4 about here>

(vii) Heterogeneous impacts depending on delay in measurement. In order to check for heterogeneous impacts depending on the considered time-frame, we conduct two tests. First, we model the relationship between delay in measurement and effect size on financial behavior

outcomes in a non-linear fashion by creating a categorical variable that distinguishes between short term (less than one month, approx. 12% of estimates), medium term (less than one year, approx. 41% of estimates), and long-term (longer than one year, approx. 47% of estimates) effects on financial behavior. Column (1) of Table S3.5 shows that short term effects tend to be higher than medium- or long-term effects on financial behavior, which is in line with the present literature (cf. Fernandes et al. 2014; Lusardi et al. 2015b). Splitting the sample according to these three time-frames, we observe that most predictors are similar in sign and magnitude in all subsamples, with some differences regarding signs and significance of predictors. It seems noteworthy, and reassuring for our results, that the subsample comprising the longer-term treatment effects appears to be driving our main results. In particular, intensity appears to matter for effect sizes to be found after a long delay between treatment and measurement. This is in line with earlier observations by Fernandes et al. (2014) that intensity may interact with delay since intervention.

<Table S3.5 about here>

(viii) Intensity. Since the intensity of financial education supports its effectiveness, we check which aspect of intensity of education drives our results. Using only the total number of hours taught as a linear predictor of effect size (and neglecting the duration of the intervention), we find that intensity does not predict effect sizes on financial behavior (available on request). This result remains the same in several variants of variable and model specifications (e.g. including polynomial forms of intensity, interaction effects between delay and intensity, and centering) and holds when effect sizes on financial literacy are regressed on this linear predictor. Thus, the intensity relative to the duration of the intervention appears to matter most for the impact on financial behavior. This finding seems to have practical implications, since it favors education with higher relative intensity, i.e. trainings with relatively more hours per week.

References in Appendix S3

Card, D., Kluve, J., and Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *Economic Journal*, 120(548): F452–F477.

Card, D., Kluve, J., and Weber, A. (2015). What works? A meta analysis of recent active labor market program evaluations. *NBER Working Paper 21431*.

Cho, Y. and Honorati, M. (2014). Entrepreneurship programs in developing countries: A meta regression analysis. *Labour Economics*, 28: 110–130.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3): 177–188.

Fernandes, D., Lynch, Jr., J.G., and Netemeyer, R.G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, 60(8): 1861–1883.

Harbord, R. M., Higgins, J. P., et al. (2008). Meta-regression in Stata. *Stata Journal*, 8(4):493–519.

Hedges, L.V., Tipton, E., and Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimate, *Research Synthesis Methods* 1(1): 39-65.

Knapp, G. and Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17): 2693–2710.

Lipsey, M.W. and Wilson, D.B. (2001). *Practical meta-analysis.* Sage, Thousand Oaks, CA.

Miller, M., Reichelstein, J., Salas, C., and Zia, B. (2015). Can you help someone become financially capable? A meta-analysis of the literature. *World Bank Research Observer*, 30(2): 220–246.

Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of Economic Perspectives*, 15(3): 131–150.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics*, 70(1):103–127.

Stanley, T. D. and Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*, Routledge, New York, NY.

Stanley, T. D. and Doucouliagos, H. (2015). Neither fixed nor random: weighted least squares meta-analysis. *Statistics in Medicine* 34(13): 2115–2127.

Tanner-Smith, E. E. and Tipton, E. (2014). Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, 5(1):13–30.

Table S3.1: Financial education treatment effect on financial behavior under various models

Outcome	(1) OLS Full pooling	(2) OLS Synthetic ES	(3) RE-Metareg	(4) RE GLS	(5) WLS 1/ SE_g	(6) Robumeta
Panel A : All						
Fin. Behavior	0.086***	0.102***	0.079***	0.093***	0.026**	0.064***
	(0.012)	(0.013)	(0.009)	(0.012)	(0.011)	(0.008)
n(Studies)	90	90	90	90	90	90
n(Effect sizes)	349	90	90	349	349	349
Panel B: RCTs						
Fin. behavior	0.082***	0.102***	0.075***	0.089***	0.067***	0.078***
	(0.014)	(0.023)	(0.013)	(0.021)	(0.013)	(0.012)
n(Studies)	40	40	40	40	40	40
n(Effect sizes	227	40	40	227	227	227
Panel C: Quasi exp.						
Fin. behavior	0.093***	0.102***	0.083***	0.100***	0.015*	0.059***
	(0.022)	(0.015)	(0.012)	(0.015)	(0.008)	(0.010)
n(Studies)	50	50	50	50	50	50
n(Effect sizes)	122	50	50	122	122	122

Notes: Column (1) shows the average effect size on fin. behavior estimated via OLS with standard errors clustered by Study ID. Column (2) shows the average effect using only one synthetic (weighted average) effect size per study. Synthetic effect sizes are estimated via within-study random effects meta-regression (DerSimonian and Laird 1986). Column (3) shows the average weighted treatment effect estimated via random effects meta-regression (DerSimonian and Laird 1986) and Knapp Hartung (2003) adjusted standard errors. The Stata command is "metareg". Column (4) shows the average treatment effect of fin. edu on fin. behavior utilizing a study random-effects GLS model. Column (5) presents results using unrestricted weighted least squares where a weight of the respective inverse standard error is assigned to each observation. Column (6) presents results from robust variance meta-regression with dependent effect size estimates (Tanner-Smith and Tipton 2014). The Stata command is "robumeta". Standard errors (clustered at the study-level for Columns (1), (4), (5), and (6)) in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level.

Table S3.2: Missing data, subsamples and giving each study equal weight

	(1) No imputations	(2) US only	(3) Classroom only	(4) Synthetic ES OLS	(5) Equal study weights
RCT	-0.052*	-0.097**	-0.080***	-0.052	-0.042
	(0.027)	(0.042)	(0.028)	(0.033)	(0.031)
TOT	0.057	0.114***	0.065**	0.107***	0.105***
	(0.041)	(0.040)	(0.028)	(0.028)	(0.035)
Delay	-0.000*	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
1/SE	0.001	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Intensity / week	0.005***	0.006*	0.004***	0.001	0.001
	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)
Duration	-0.000	-0.001	-0.000	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Low income clients	-0.047**	-0.003	-0.054***	-0.043	-0.049**
	(0.020)	(0.025)	(0.017)	(0.027)	(0.022)
Years of schooling	-0.022***		-0.021***	-0.022**	-0.020**
	(0.007)		(0.006)	(0.009)	(0.009)
Low/lower-mid .econ	-0.113**		-0.108**	-0.113*	-0.108*
	(0.044)		(0.041)	(0.061)	(0.059)
Mandatory	-0.086*	-0.097***	-0.043	-0.097**	-0.095***
	(0.049)	(0.033)	(0.028)	(0.038)	(0.029)
Teachable moment	0.058	0.129***	0.075**	0.058**	0.058**
	(0.052)	(0.035)	(0.033)	(0.028)	(0.026)
Constant	0.359***	0.042	0.364***	0.364***	0.344***
	(0.097)	(0.031)	(0.095)	(0.118)	(0.119)
R^2	0.125	0.340	0.177	0.297	0.206
n (Studies)	35	55	70	90	90
n (Effect sizes)	24	135	317	90	349

Notes: Column (1) reports results for complete cases only. Columns (2) present results for the sample split of USA studies only. These splits include only variables for which differential information from at least two studies are available. Column (3) presents results using one synthetic effect size (weighted within-study average effect size across all outcomes) per study. Column (4) shows results by weighting each observation by the inverse number of observations of the study the observation is nested in. Standard errors (clustered at the study-level for all Columns but (4)) in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level.

Table S3.3: Alternative meta-regression models

	(1) Probit 5%	(2) Ordered probit 10%	(3) RE GLS	(4) WLS 1/SE(g) weights	(5) WLS 1/Var(g) weights
RCT	-0.794***	-0.802***	-0.087***	-0.086***	-0.044**
	(0.225)	(0.196)	(0.024)	(0.020)	(0.022)
TOT	0.052	0.002	0.049**	0.038**	0.058***
	(0.189)	(0.176)	(0.023)	(0.016)	(0.015)
Delay	-0.001**	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
1/SE	0.001	-0.000			
	(0.001)	(0.001)			
SE_g			0.486***	0.611**	
			(0.173)	(0.272)	
SE_g^2					3.147**
					(1.496)
Intensity /week	0.018	0.027*	0.003**	0.003*	0.006**
	(0.014)	(0.015)	(0.002)	(0.002)	(0.003)
Duration	0.008*	-0.000	0.000	0.000	0.000
	(0.004)	(0.005)	(0.001)	(0.000)	(0.000)
Low inc. clients	-0.566***	-0.561***	-0.060***	-0.014*	-0.000
	(0.160)	(0.148)	(0.019)	(0.007)	(0.002)
Years of schooling	-0.154***	-0.136***	-0.024***	-0.022***	-0.018***
	(0.058)	(0.044)	(0.006)	(0.005)	(0.006)
Low/lower-mid. econ.	-0.872**	-0.792**	-0.105**	-0.086***	-0.076*
	(0.392)	(0.314)	(0.042)	(0.032)	(0.045)
Mandatory	0.172	0.130	-0.030	-0.026	-0.017
	(0.245)	(0.272)	(0.026)	(0.020)	(0.018)
Teach. moment	0.326	0.404**	0.063***	0.042**	0.068***
	(0.219)	(0.192)	(0.024)	(0.017)	(0.015)
Constant cut 1		-3.977***			
		(0.636)			
Constant cut 2		-1.999***			
		(0.594)			
Constant	2.009**		0.356***	0.304***	0.210***
	(0.783)		(0.079)	(0.066)	(0.079)
R^2			0.197	0.301	0.336
Pseudo R^2	0.109	0.084			
n (Studies)	90	90	90	90	90
n (Effect Sizes)	349	349	349	349	349

Notes: Dependent variable in columns (1) and (2) is a categorical indicator of sign and significance of intervention impact. Dependent variable in columns (3) and (4) is effect size (Hedges' g) on financial behavior. Column (1) reports results from probit-regression with a binary outcome indicating whether financial education had a significantly positive effect on financial behavior at the 5%-level. Column (2) provides results for ordered probit regression with a dependent categorical variable taking the value "-1" if financial education had a significantly negative impact on financial behavior, "0" if financial education had an insignificant effect on financial behavior, and "1" if financial education had a significant positive effect on financial behavior at the 10%-level. Column (3) reports results from GLS random-effects regression. Column (4) reports results of weighted least squares estimation with inverse variance weights. Standard errors clustered at the study-level in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level.

Table S3.4: Random effects meta-regression on synthetic effect sizes with inverse variance weights

	(1) MM	(2) MM	(3) REML	(4) REML	(5) EB	(6) EB
RCT	-0.066***	-0.066***	-0.065***	-0.065***	-0.066***	-0.066***
	(0.021)	(0.020)	(0.021)	(0.020)	(0.022)	(0.022)
TOT	0.061***	0.061***	0.061***	0.061***	0.063***	0.063***
	(0.020)	(0.019)	(0.020)	(0.019)	(0.020)	(0.020)
Delay	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Intensity /week	0.002	0.002	0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Duration	0.000	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Low inc. clients	-0.024	-0.024	-0.024	-0.024	-0.025	-0.025
	(0.016)	(0.016)	(0.016)	(0.015)	(0.017)	(0.017)
Years of schooling	-0.015**	-0.015***	-0.014**	-0.014***	-0.015**	-0.015**
	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)
Low/lower inc. econ.	-0.044	-0.044	-0.043	-0.043	-0.046	-0.046
	(0.039)	(0.038)	(0.039)	(0.037)	(0.040)	(0.040)
Mandatory	-0.052**	-0.052***	-0.051**	-0.051***	-0.053**	-0.053**
	(0.020)	(0.019)	(0.020)	(0.019)	(0.021)	(0.021)
Teach. moment	0.053***	0.053***	0.052***	0.052***	0.053***	0.053***
	(0.018)	(0.017)	(0.018)	(0.017)	(0.018)	(0.018)
Constant	0.251***	0.251***	0.249***	0.249***	0.256***	0.256***
	(0.075)	(0.073)	(0.075)	(0.071)	(0.076)	(0.076)
I^2	81.14%	81.14%	81.14%	81.14%	81.14%	81.14%
Adj. R^2	-	-	0.442	0.442	0.474	0.474
n (Studies)	90	90	90	90	90	90
n (Effect Sizes)	90	90	90	90	90	90
Adjusted errors	yes	no	yes	no	yes	no

Notes: Results from random-effects meta-regression (DerSimonian and Larid 1986) with and without Knapp and Hartung (2003) adjusted standard errors, respectively. Dependent variable is effect size (Hedges' g) on financial behavior weighted by its inverse variance. Columns (1) and (2) show results for method of moments (MM) estimates. Columns (3) and (4) show results for restricted maximum likelihood (REML) estimates. Columns (4) and (5) show results from empirical bayes estimates. The Stata command is *metareg* (Hardbord and Higgins 2008). Standard errors in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level.

Table S3.5: Effect sizes on financial behavior and heterogeneity of treatment effects by delay in measurement of treatment effects

	(1) Financial behavior	(2) Short term subsample	(3) Medium term subsample	(4) Long term subsample
RCT	-0.061**	0.148	-0.085***	-0.073*
	(0.026)	(0.102)	(0.027)	(0.038)
TOT	0.043*	-0.221**	0.043	0.062
	(0.025)	(0.078)	(0.032)	(0.049)
Short term	0.089**			
	(0.039)			
Medium term	-0.006			
	(0.018)			
1/SE	-0.000	-0.005**	-0.000	0.000
	(0.000)	(0.002)	(0.000)	(0.000)
Intensity /week	0.004***	0.006	0.002	0.004***
	(0.001)	(0.007)	(0.002)	(0.001)
Duration	-0.000	0.010**	0.000	0.000
	(0.001)	(0.005)	(0.001)	(0.001)
Low inc. clients	-0.044***	-0.046	-0.041**	-0.045**
	(0.014)	(0.087)	(0.020)	(0.019)
Years of schooling	-0.021***	-0.103**	-0.011	-0.021**
	(0.005)	(0.047)	(0.008)	(0.009)
Low/lower inc. econ.	-0.122***	-1.127***	0.034	-0.156***
	(0.041)	(0.318)	(0.055)	(0.058)
Mandatory	-0.041**	-0.076	0.003	-0.056***
	(0.019)	(0.097)	(0.047)	(0.021)
Teach. moment	0.090***	0.202*	0.009	0.109***
	(0.028)	(0.108)	(0.032)	(0.024)
Constant	0.332***	1.634**	0.235**	0.332***
	(0.077)	(0.624)	(0.101)	(0.119)
R^2	0.204	0.457	0.073	0.319
n (Studies)	90	18	24	53
n (Effect Sizes)	349	42	143	164

Notes: Results from OLS meta-regression with robust standard errors clustered at the study-level. Dependent variable is effect size (Hedges' g) on financial behavior. Standard errors in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level.

Appendix S4: Publication bias and heterogeneity of study quality

We show examinations of conventional visual tests for publication bias in order to address the so-called file drawer problem (cf. Stanley and Doucouliagos 2012, p. 73) and examine the sample of studies for heterogeneous results depending on study quality. Note that we also use formal econometric methods, (i.e. alternative regression approaches) in Appendix S3 that are in principle capable of generating unbiased estimates in the presence of publication selection (see table S3.4 columns 4 and 5).

Publication bias. We conduct visual tests for overall publication bias (funnel asymmetry), so-called funnel plots (cf. Figures S4.1 and S4.2). Precision of the estimated treatment effect should increase in larger studies. Thus, we scatter effect sizes (multiple effects per study) against the standard errors of the effect size estimates (inverted y-axis). Effect estimates from small studies (larger sampling errors) should scatter more widely at the bottom of the graph, with the spread decreasing as standard errors decrease. In the absence of bias, the plot resembles a symmetrical inverted funnel. Therefore, asymmetry indicates a publication bias in the sense that negative or non-results are under-represented (i.e. not published at all). Inspecting the two plots indicates that symmetry is higher for effect sizes on financial behavior than for effect sizes on literacy but both outcomes may be affected by publication biases in the sense that the overall treatment effect may suffer from a slight upward bias. This conclusion, however, requires the assumption that non-results are not published at all (i.e. the file drawer problem).

<Figure S4.1 and Figure S4.2 about here>

This assumption may be more plausible for quasi- and natural experiments than for RCTs, as results from rigorous randomized experiments are likely to be published irrespective of their results. Therefore, we perform the same visual check on the subsample of RCTs only (cf. Figures S4.3 and S4.4). Indeed, these plots appear to be more symmetric indicating that

publication bias may primarily be an issue within the sample of non-randomized studies. As (i) nearly 40 percent of our sample is comprised of RCTs; (ii) we control for research design in all our regressions; and (iii) our main results replicate within the subsample of RCTs, we suggest that publication biases are not an issue for our analysis. However, we also test the robustness of our results using weighted least squares and controlling for the standard error (or the squared standard error, i.e. the variance of the estimate), which is advocated as a robust method in the presence of publication selection (cf. Stanley and Doucouliagos 2012).

<Figure S4.3 and Figure S4.4 about here>

Publication status and quality. Another concern in any meta-analysis is the issue of biases arising from the aggregation of results from studies with different publication status and quality. On the one hand, researchers fear that the tendency of the scientific community to favor statistically significant positive results over insignificant non-results may lead to biased estimates favoring the rejection of the null hypothesis of a zero-effect of financial education on relevant outcomes. The standard solution in the meta-analysis literature is to include as many unpublished studies (grey literature) as possible to address this potential source of bias *a priori*.

On the other hand, economists fear that by aggregating studies of different publication status and quality, the results suffer due to the lack of empirical rigor in grey-literature primary studies. To shed light on this issue in the financial education literature, we compare average effect sizes of financial education interventions by different types of publication status and indicators of quality. Table S4.1 compares average effect sizes on financial literacy and behavior by publication status in an academic journal. Interestingly, a bias affects only the effect size estimates on financial literacy, as they appear to be more than twice as high in published than in unpublished papers (t=3.863). Turning to effect sizes on financial behavior, however, we observe no significant difference in average effect sizes between published and unpublished studies.

<Table S4.1 about here>

Considering indicators of study quality, we code the article influence score (ISI web of knowledge) of the respective journal (and year) for every publication and assign a value of 0 for studies available as working papers. Comparing influential (article influence score >1) with less influential (≤1) publications, we find that the quality bias for financial literacy is now insignificant (t=0.328): Moreover, influential journals tend to publish studies with 0.04 standard deviation units smaller effect sizes on behavior (t=-2.189) than non-influential journals. Thus, more rigorous work reports a slightly smaller average treatment effect than presumably less rigorous work.

Next, we code the number of citations for each publication as reported in Google Scholar (as of October 31, 2016). The mean number of citations per article is 53.91and we split the sample in studies cited above and below this threshold value. Again, we find no significant differences between highly cited studies and others: If anything, highly cited studies tend to report smaller average effect sizes on financial behavior than studies with few citations. Overall, we see that quality bias appears to be not an issue that alters the conclusions in this literature concerned with effects on financial behavior.

Reference in Appendix S4

Stanley, T. D. and Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*, Routledge, New York, NY.

Table S4.1: Effect sizes by publication status and indicators of publication quality

Outcome	Status / Quality	Studies	Obs.	ES (g)	SE_g	p-value	Diff. (t-value)
Fin. Literacy	Published	36	106	0.343	0.066	0.000	0.179***
	Unpublished	31	85	0.164	0.039	0.000	(3.863)
Fin. Behavior	Published	50	142	0.087	0.019	0.000	0.004
	Unpublished	40	200	0.083	0.016	0.000	(0.211)
Fin. Literacy	High influence	11	36	0.247	0.028	0.000	0.020
	Low influence	56	155	0.267	0.043	0.000	(0.328)
Fin. Behavior	High influence	27	90	0.053	0.020	0.013	-0.043**
	Low influence	63	252	0.096	0.015	0.000	(-2.189)
Fin. Literacy	Highly cited	10	17	0.249	0.068	0.005	-0.016
	Few citations	57	174	0.265	0.045	0.000	(-0.195)
Fin. Behavior	Highly cited	37	73	0.070	0.024	0.006	-0.018
	Few citations	53	269	0.089	0.014	0.000	(-0.879)

Notes: ES(g) and SE_g are results from an unweighted OLS regression with standard errors clustered by study ID. Samples are split by an indicator of publication in an academic journal (published / unpublished), an indicator of high and low influence (article influence score >1), and an indicator of highly cited articles (Google scholar citations > mean(citations)).

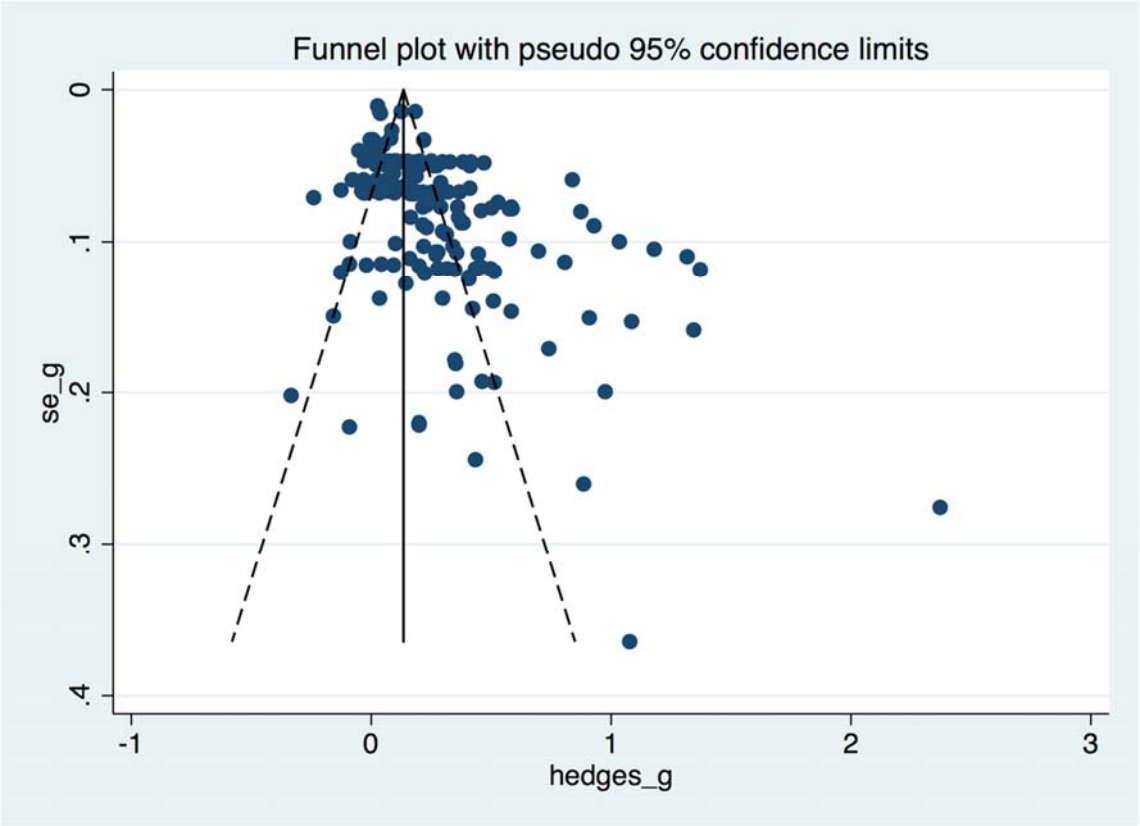


Figure S4.1: Funnel plot of treatment effects on financial literacy

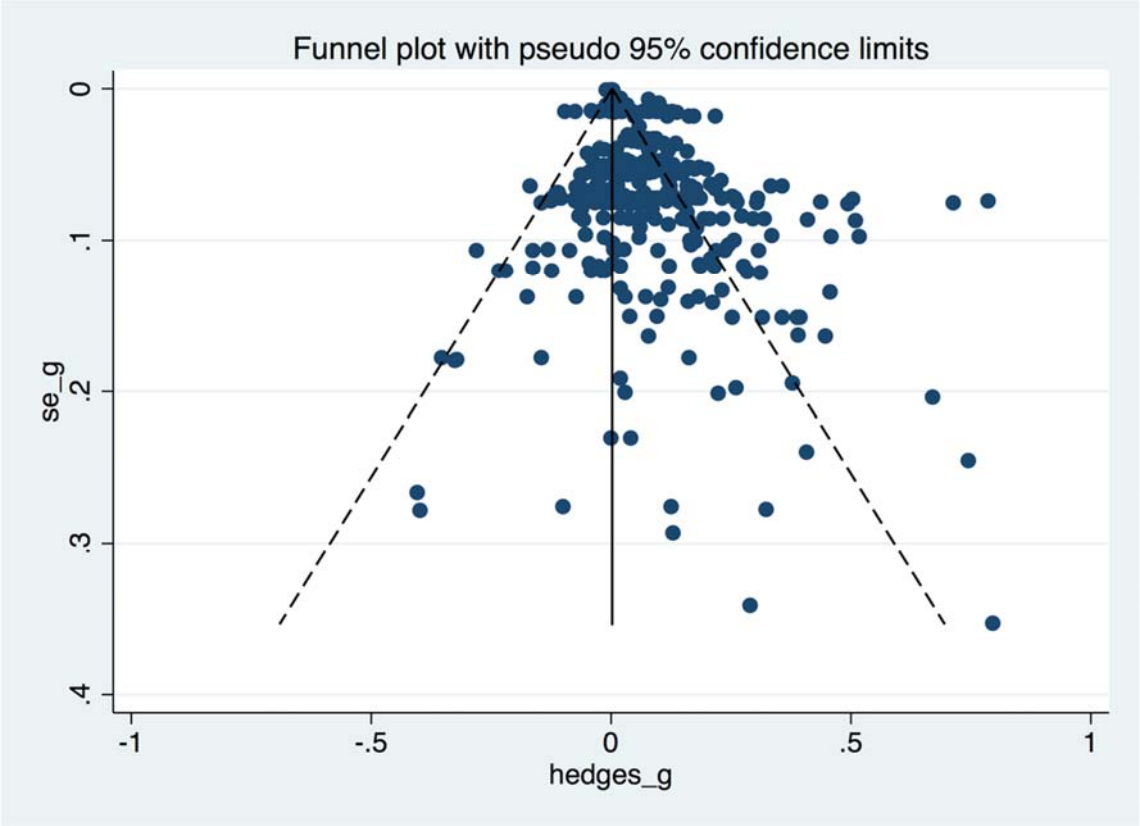


Figure S4.2: Funnel plot of treatment effects on financial behavior

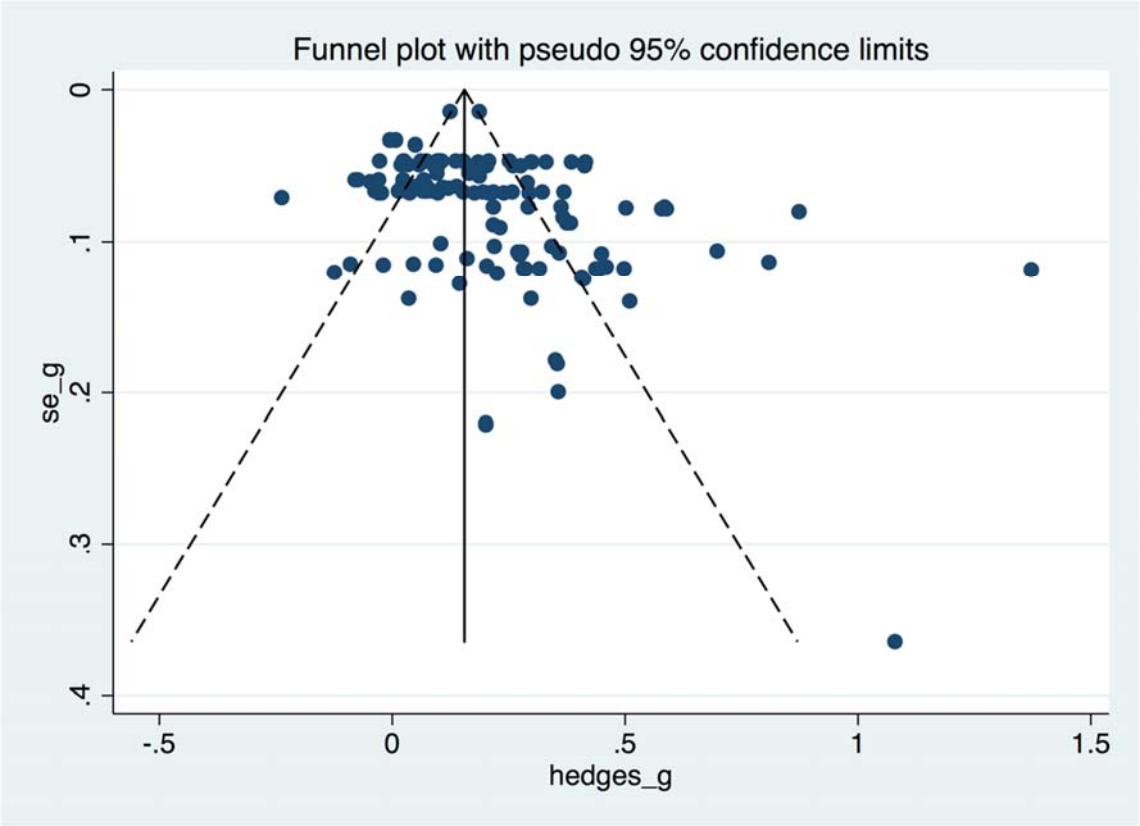


Figure S4.3: Funnel plot of treatment effects on financial literacy within the subsample of RCTs only

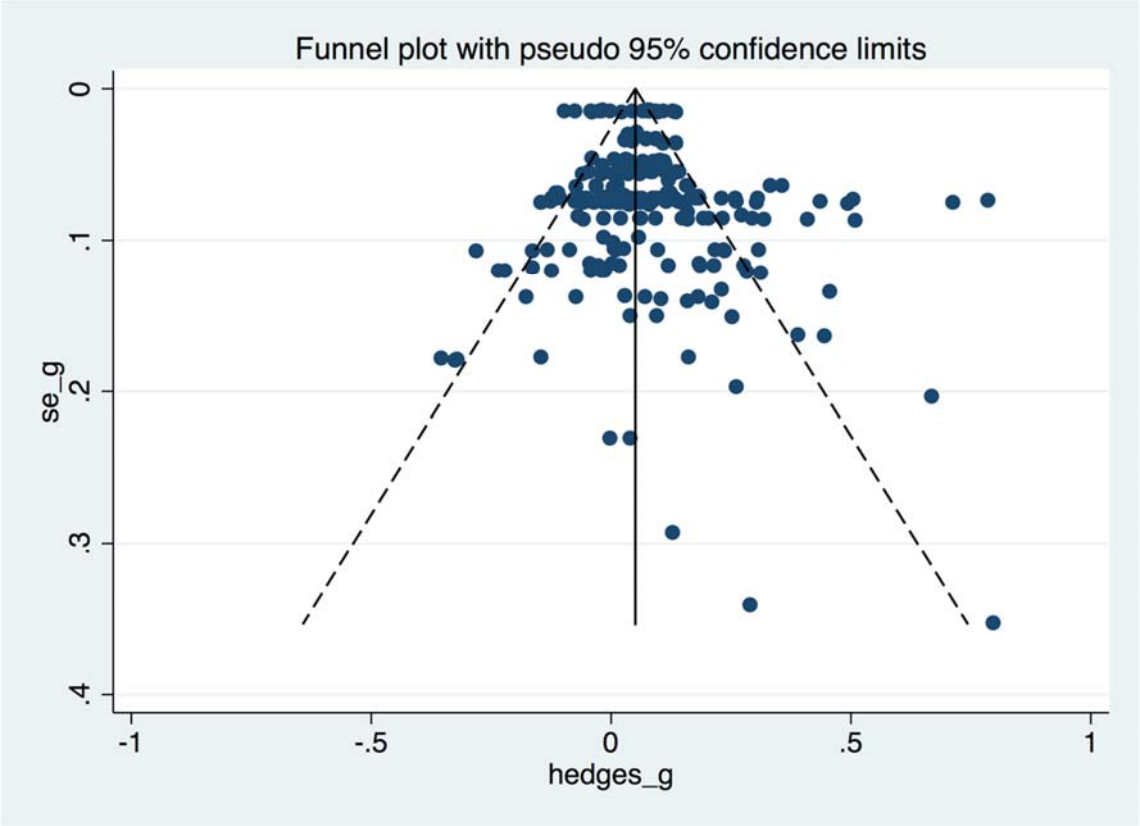


Figure S4.4: Funnel plot of treatment effects on financial behavior within the subsample of RCTs only

Appendix S5: Overview of studies included in the statistical meta-analysis

Table S5.1: Overview of financial education studies included in our analysis

Study	Country	Research design	Target group		Intervention	
			Mean age	*Low-income*	*Channel*	*Teach. moment*
Agarwal et al. 2009	USA	Natural exp.	-	Yes	Counseling	Yes
Agarwal et al. 2010	USA	Natural exp.	-	Yes	Counseling	Yes
Ambuhel et al. 2014	USA	RCT	29	Yes	Online	No
Asarta et al. 2014	USA	Quasi exp.	15	-	Classroom	No
Barcellos et al. 2012	USA	RCT	52	No	Online	No
Baron-Donovan et al. 2005	USA	Quasi exp.	44	No	Classroom	No
Barua et al. 2012	Singapore	RCT	37	Yes	Classroom	Yes
Batty et al. 2015	USA	RCT	9	Yes	Classroom	No
Bauer et al. 2011	USA	Quasi exp.	-	Yes	Classroom	No
Bayer et al. 2009	USA	Natural exp.	-	No	Classroom	Yes
Becchetti et al. 2013	Italy	RCT	18	-	Classroom	No
Berg and Zia 2013	South Africa	RCT	20	Yes	Mass Media	No
Bell et al. 2009	USA	Quasi exp.	22	No	Classroom	No
Bernheim and Garrett 2003	USA	Natural exp.	39	No	Classroom	Yes
Bernheim et al. 2001	USA	Natural exp.	40	No	Classroom	No
Berry et al. 2015	Ghana	RCT	11	-	Classroom	Yes
Bjorvatn and Tungodden 2010	Tanzania	RCT	39	-	Classroom	Yes
Brown et al. 2016	USA	Natural exp.	28	-	Classroom	No
Brugiavini et al. 2015	Italy	RCT	23	No	Classroom	No
Bruhn and Zia 2013	Bosnia and Herzegovina	RCT	28	Yes	Classroom	Yes
Bruhn et al. 2014	Mexico	RCT	33	-	Classroom	No
Bruhn et al. 2013	Brazil	RCT	16	Yes	Classroom	No
Butt et al. 2008	USA	Quasi exp.	12	No	Classroom	No
Calderone et al. 2013	India	RCT	45	Yes	Classroom (video)	Yes

Carlin and Robinson 2012	USA	Quasi exp.	16	No	Classroom	No
Carpena et al. 2011	India	RCT	39	Yes	Classroom	Yes
Carpena et al. 2015	India	RCT	39	Yes	Classroom + Counseling	Yes
Chen and Heath 2012	USA	Quasi exp.	9	-	Classroom	No
Choi et al. 2005	USA	Natural exp.	40	No	Classroom	No
Choi et al. 2010	USA	RCT	31	No	Info. nudge	No
Choi et al. 2011	USA	Natural exp.	64	No	Info. nudge	No
Clancy et al. 2001	USA	Natural exp.	36	Yes	Classroom	Yes
Clark et al. 2006	USA	Quasi exp.	54	No	Classroom	Yes
Clark et al. 2015	USA	Quasi exp.	44	No	Online	No
Clark et al. 2014	USA	RCT	35	No	Info. nudge	Yes
Clark et al. 2010	USA	Quasi exp-	57	No	Classroom	Yes
Cole and Shastry 2010	USA	Natural exp.	-	No	Classroom	No
Cole et al. 2013	India	RCT	48	Yes	Counseling	Yes
Cole et al. 2014	USA	Natural exp.	17	Yes	Classroom	No
Cole et al. 2011	Indonesia	RCT	41	Yes	Classroom	Yes
Collins 2013	USA	RCT	39	Yes	Classroom	No
Custers 2011	India	RCT	34	Yes	Classroom	Yes
Danes and Haberman 2004	USA	Quasi exp.	15	No	Classroom	No
Danes et al. 1999	USA	USA	15	No	Classroom	No
De Mel et al. 2011	Sri Lanka	Quasi exp.	41	-	Classroom	Yes
DeLaune et al. 2010	USA	Quasi exp.	18	No	Classroom	No
Ding et al. 2008	USA	Natural exp.	-	Yes	Counseling	Yes
Doi et al. 2014	Indonesia	RCT	44	Yes	Classroom	Yes
Dolvin and Templeton 2006	USA	Quasi exp.	46	No	Classroom	Yes
Drexler et al. 2014	Dominican Republic	RCT	41	Yes	Classroom	Yes
Duflo and Saez 2003	USA	RCT	38	No	Info. nudge	Yes

Elliehausen et al. 2007	USA	Natural exp.	41	No	Counseling	No
ETI 2008	USA	Quasi exp.	14	-	Classroom	No
Field et al. 2010	India	RCT	32	Yes	Classroom	Yes
Fort et al. 2016	Italy	Natural exp.	-	-	Info. Nudge	No
Garman et al. 1999	USA	Quasi exp.	43	No	Classroom	Yes
Gaurav et al. 2011	India	RCT	50	Yes	Classroom	Yes
Gibson et al. 2014	New Zealand / Australia	RCT	-	Yes	Classroom	Yes
Gill and Bhattacharya 2015	USA	Quasi exp.	17	Yes	Classroom	No
Gine and Mansuri 2014	Pakistan	RCT	38	Yes	Classroom	Yes
Gine et al. 2013	Kenya	RCT	49	Yes	Edu. materials	Yes
Go et al. 2012	USA	Quasi exp.	9	Yes	Classroom	No
Goda et al. 2014	USA	Quasi exp.	45	No	Info. nudge	No
Goldsmith and Goldsmith 2006	USA	Quasi exp.	19	No	Classroom	No
Grimes et al. 2010	USA	Natural exp.	51	No	Classroom	No
Grinstein-Weiss et al. 2015	USA	Natural exp.	36	Yes	Classroom	Yes
Han et al. 2009	USA	RCT	41	Yes	Classroom	Yes
Hartaska and Gonzalez-Vega 2005	USA	Natural exp.	-	Yes	Counseling	Yes
Hartaska and Gonzalez-Vega 2006	USA	Natural exp.	35	No	Counseling	Yes
Harter and Harter 2009	USA	Quasi exp.	-	Yes	Classroom	No
Harter and Harter 2010	USA	Quasi exp.	17	No	Classroom	No
Haynes et al- 2011	USA	RCT	55	Yes	Online	No
Haynes-Bordas et al. 2008	USA	Quasi exp.	38	Yes	Classroom	Yes
Heinberg et al. 2014	USA	RCT	35	No	Online	No
Hershey et al. 2003	USA	RCT	34	Yes	Classroom	No

Hirad and Zorn 2001	USA	Natural exp.	-	Yes	Mixed	Yes
Hospido et al. 2015	Spain	Quasi exp.	15	-	Classroom	No
Jamison et al. 2014	Uganda	RCT	24	No	Classroom	Yes
Kimball and Shumway 2010	USA	Natural exp.	50	No	Mixed	Yes
Krause et al. 2016	Tanzania	Quasi exp.	-	-	Classroom	Yes
Loke et al. 2015	USA	Quasi exp.	15	Yes	Classroom	Yes
Lusardi 2002	USA	Natural exp.	-	-	Classroom	Yes
Lusardi 2005	USA	Natural exp.	55	No	Classroom	No
Lusardi and Mitchell 2007	USA	Natural exp.	53	No	Classroom	No
Lusardi et al. 2014	USA	RCT	50	No	Online	No
Lührmann et al. 2015	Germany	Quasi exp.	14	Yes	Classroom	No
Maki 2004	USA	Natural exp.	40	No	Classroom	No
Mandell 2006	USA	Quasi exp.	12	-	Classroom	No
Mandell 2009a	USA	Quasi exp.	-	-	Classroom	No
Mandell 2009b	USA	Quasi exp.	13	-	Classroom	No
Mandell and Schmid-Klein 2009	USA	Quasi exp.	16	-	Classroom	No
Mills et al. 2004	USA	RCT	36	Yes	Classroom	No
Muller 2003	USA	Natural exp.	-	No	Classroom	No
Pang 2010	Hong Kong, China	Quasi exp.	19	-	Classroom	No
Peng et al. 2010	USA	Natural exp.	35	No	Classroom	Yes
Quercia and Spader 2008	USA	Natural exp.	30	Yes	Classroom	No
Reich and Berman 2015	USA	RCT	30	Yes	Classroom	Yes
Romagnoli and Trifildis 2013	Italy	Quasi exp.	14	No	Classroom	No
Sanders et al. 2007	USA	Quasi exp.	35	Yes	Classroom	Yes
Sarr et al. 2012	India	RCT	38	Yes	Classroom	Yes
Sayinzoga et al. 2016	Rwanda	RCT	40	Yes	Classroom	Yes
Schreiner et al. 2001	USA	Natural exp.	-	Yes	Classroom	Yes

Seshan and Yang 2014	Qatar	RCT	40	Yes	Classroom	Yes
Skimmyhorn 2016	USA	Natural exp.	21	Yes	Classroom	No
Skimmyhorn et al. 2016	USA	RCT	-	-	Classroom	No
Supanantaroek et al. 2016	Uganda	RCT	-	-	Classroom	Yes
Song 2012	China	RCT	45	No	Info. nudge	No
Tennyson and Nguyen 2001	USA	Natural exp.	17	Yes	Classroom	No
Vacroe et al. 2005	USA	Quasi exp.	17	-	Classroom	No
Walstad et al. 2010	USA	Quasi exp.	18	No	Classroom	No
Wiener et al. 2005	USA	Quasi exp.	39	No	Classroom	Yes
Xiao et al. 2012	USA	Natural exp.	18	No	Classroom	No
Yetter and Suiter 2015	USA	RCT	24	Yes	Classroom	No

Appendix S6: References for studies included in the statistical meta-analysis

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., and Evanoff, D.D. (2009). Do financial counseling mandates improve mortgage choice and performance? Evidence from a natural experiment. *Federal Reserve Bank of Chicago Working Paper* 2009–07.

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., and Evanoff, D.D. (2010). Learning to cope: Voluntary financial education and loan performance during a housing crisis. *American Economic Review: Papers and Proceedings* 100: 495–500.

Ambuehl, S., Bernheim, B.D., and Lusardi, L. (2014). The effect of financial education on the quality of decision making. *NBER Working Paper 20618.*

Asarta, C.J., Hill, A.T., and Meszaros, B.T. (2014). The features and effectiveness of the keys to financial success curriculum. *International Review of Economics Education*, 16, Part A(0): 39–50.

Barcellos, S.H., Smith, J.P., Yoong, J.K., and Carvalho, L. (2012). Barriers to immigrant use of financial services. The role of language skills, U.S. experience, and return migration expectations. *Financial Literacy Center Working Paper*, *Wr-923-SSA. Dartmouth College and the Wharton School.*

Baron-Donovan, C., Wiener, R.L., Gross, K., and Block-Lieb, S. (2005). Financial literacy teacher training: A multiple-measure evaluation. *Journal of Financial Counseling and Planning*, 16(2): 63–75.

Barua, R., Shastry, G.K., and Yang, D. (2012). Evaluating the effect of peer-based financial education on savings and remittances for foreign domestic workers in Singapore. *Working Paper. Singapoore Management University, Wellesley College, and University of Michigan.*

Batty, M., Collins, J.M., and Odders-White, E. (2015). Experimental evidence on the effects of financial education on elementary school students' knowledge, behavior, and attitudes. *Journal of Consumer Affairs*, 49(1): 69–96.

Bauer, J.W., Son, S., Hur, J., Anderson-Porich, S., Heins, R.H., Petersen, C., Hooper, S., Marczak, M., Olson, P.D., and Wiik, N.B. (2011). Dollar works 2: Impact evaluation report. *University of Minnesota, Extension. St Paul, MN.*

Bayer, P.J., Bernheim, B.D., and Scholz, J.K. (2009). The effects of financial education in the workplace: Evidence from a survey of employers. *Economic Inquiry*, 47(4): 605–624.

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

Berg, G. and Zia, B. (2013). Harnessing emotional connections to improve financial decisions. Evaluating the impact of financial education in mainstream media. *World Bank Policy Research Working Paper 6407.*

Bell, C., Gorin, D., and Hogarth, J.M. (2009). Does financial education affect soldiers' financial behaviors? *Networks Financial Institute Working Paper 2009-WP-08, Indiana State University.*

Bernheim, B.D.. and Garrett, D.M. (2003). The effects of financial education in the workplace: Evidence from a survey of households. *Journal of Public Economics*, 87(7–8): 1487–1519.

Bernheim, B.D., Garrett, D.M., and Maki, D.M. (2001). Education and saving: the long-term effects of high school financial curriculum mandates. *Journal of Public Economics*, 80(3): 435–465.

Berry, J., Karlan, D., and Pradhan, M. (2015). The impact of financial education for youth in Ghana. *NBER Working Paper 21068*.

Bhattacharya, R., Gill, A., and Stanley, D. (2016). The effectiveness of financial literacy instruction: The role of individual development accounts participation and the intensity of instruction. *Journal of Financial Counseling and Planning*, 27(1): 20–35.

Bjorvatn, K, and Tungodden, B. (2010). Teaching business in Tanzania: Evaluating participation and performance. *Journal of the European Economic Association*, 8 (2-3): 561–570.

Brown, M., Grigsby, J., van der Klaauw, W., Wen, J., and Zafar, B. (2016). Financial education and the debt behavior of the young. *Review of Financial Studies*, 29(9): 2490–2522.

Brugiavini, A., Cavapozzi, D., Padula, M., and Pettinicchi, Y. (2015). Financial education, literacy and investment attitudes. *SAFE Working Paper No. 86. University of Venice and SAFE-Center, University of Frankfurt.*

Bruhn, M. and Zia, B. (2013). Stimulating managerial capital in emerging markets: The impact of business training for young entrepreneurs. *Journal of Development Effectiveness*, 5(2): 232–266.

Bruhn, M., Ibarra, G.L. and McKenzie, D. (2014). The minimal impact of a large-scale financial education program in Mexico city. *Journal of Development Economics*, 108: 184–189.

Bruhn, M., Le˜ao, L. d. S., Legovini, A., Marchetti, R., and 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): 256–95.

Butt, N.M., Haessler, S.J., and Schug, M.C. (2008). An incentives-based approach to implementing financial fitness for life in the Milwaukee public schools. *Journal of Private Enterprise*, 24(Fall 2008): 165–173.

Calderone, M., Fiala, N., Mulaj, F. Sadhu, S., and Sarr, L. (2013). When can financial education affect savings behavior?⊠Evidence from a randomized experiment among low income clients of branchless banking in India. *Working Paper, German Institute for Economic Research, Berlin and World Bank, Washington D.C.*

Carlin, B.I. and Robinson, D.T. (2012). What does financial literacy training teach us? *Journal of Economic Education*, 43(3): 235–247.

Carpena, F., Cole, S., Shapiro, J., and Zia, B. (2011). Unpacking the causal chain of financial literacy. *World Bank Policy Research Working Paper 5798.*

Carpena, F., Cole, S., Shapiro, J., and Zia, B (2015). The ABCs of financial education. Experimental evidence on attitudes, behavior, and cognitive biases. *World Bank Policy Research Working Paper 7413.*

Carter, M.R., Laajaj, R., and Yang, D. (2016). Savings, Subsidies, and Technology Adoption: Field Experimental Evidence from Mozambique. *Unpublished working paper.*

Chen, W. and Heath, J.A. (2012). The efficacy of financial education in the early grades. In Laney, J.D. and Lucey, T.A., editors, *Reframing Financial Literacy: Exploring the Value of Social Currency*, 189–207. Information Age Publishing. Charlotte, N.C.

Choi, J.J., Laibson, D., and Madrian, B.C. (2005). Are empowerment and education enough? Underdiversification in 401 (k) plans. *Brookings Papers on Economic Activity*, 2005(2): 151–213.

Choi, J.J., Laibson, D., and Madrian, B.C. (2010). Why does the law of one price fail? An experiment on index mutual funds. *Review of Financial Studies*, 23(4): 1405–1432.

Choi, J.J., Laibson, D., and Madrian, B.C. (2011). $100 bills on the sidewalk: Suboptimal investment in 401 (k) plans. *Review of Economics and Statistics*, 93(3): 748–763.

Clancy, M., Grinstein-Weiss, M., and Schreiner, M. (2001). Financial education and savings outcomes in individual development accounts. *Working Paper, Washington University in St. Louis Center for Social Development, St. Louis, MO.*

Clark, R.L., d'Ambrosio, M.B., McDermed, A.A., and Sawant, K. (2006). Retirement plans and saving decisions: The role of information and education. *Journal of Pension Economics and Finance*, 5: 45–67.

Clark, R., Lusardi, A., and Mitchell, O.S. (2015). Employee financial literacy and retirement plan behavior: A case study. *NBER Working Paper 21461.*

Clark, R.L., Maki, J.A., and Morrill, M.S. (2014). Can simple informational nudges increase employee participation in a 401(k) plan? *Southern Economic Journal*, 80(3): 677–701.

Clark, R.L., Morrill, M.S., and Allen, S.G. (2010). Employer-provided retirement planning programs. In Clark, R.L. and Mitchell, O.S., editors, *Reorienting Retirement Risk Management*, 36–64. Oxford University Press.

Cole, S. and Shastry, G.K. (2010). Is high school the right time to teach savings behavior? The effect of financial education and mathematics courses on savings. *Harvard Business School Working Paper.*

Cole, S., Gine, X., Tobacman, J., Topalova, P., Townsend, R., and Vickery, J. (2013). Barriers to household risk management: Evidence from India. *American Economic Journal: Applied Economics*, 5(1): 104–135.

Cole, S., Paulson, A., and Shastry, G.K. (2014), High school curriculum and financial outcomes: The impact of mandated personal finance and mathematics courses. *Harvard Business School Working Paper 13-064.*

Cole, S., Sampson, T., and Zia, B. (2011). Prices or knowledge? What drives demand for financial services in emerging markets? *Journal of Finance*, 66(6): 1933–1967.

Collins, J.M. (2013). The impacts of mandatory financial education: Evidence from a randomized field study. *Journal of Economic Behavior and Organization*, 95: 146–158.

Custers, A. (2011). Furthering financial literacy: Experimental evidence from a financial literacy program for microfinance clients in Bhopal, India. *LSE International Development Working Paper 11-113*, London.

Danes, S.M. and Haberman, H. (2004). Evaluation of the NEFE high school financial planning program: 2003-2004. *National Endowment for Financial Education Project Report, University of Minnesota.*

Danes, S.M, Huddleston-Casas, C., and Boyce, L. (1999). Financial planning curriculum for teens: Impact evaluation. *Financial Counseling and Planning*, 10(1): 26-39.

De Mel, S., McKenzie, D., and Woodruff, C. (2011). Getting credit to high return microentrepreneurs: The results of an information intervention. *World Bank Economic Review*, 25(3): 456–485.

DeLaune, L.D., Rakow, J.S., and Rakow, K. (2010). Teaching financial literacy in a co-curricular service-learning model. *Journal of Accounting Education*, 28(2): 103–113.

Ding, L., Quercia, R.G., and Ratcliffe, J. (2008). Post-purchase counseling and default resolutions among low- and moderate-income borrowers. *Journal of Real Estate Research*, 30(3): 315–344.

Doi, Y., McKenzie, D., and Zia, B. (2014). Who you train matters: Identifying combined effects of financial education on migrant households. *Journal of Development Economics*, 109(0): 39–55.

Dolvin, S. and Templeton, W.K. (2006), Financial education and asset allocation. *Financial Services Review*, 15: 133–149.

Drexler, A., Fischer, G., and Schoar, A. (2014). Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, 6(2): 1–31.

Duflo, E. and Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Quarterly Journal of Economics*, 118(3): 815–842.

Elbogen, E. B., Hamer, R. M., Swanson, J. W., and Swartz, M. S. (2016). A randomized clinical trial of a money management intervention for veterans with psychiatric disabilities. *Psychiatric Services*, 0(0):appi.ps.201500203. PMID: 27181733.

Elliehausen, G., Lundquist, C.E., and Staten, M.E. (2007). The impact of credit counseling on subsequent borrower behavior. *Journal of Consumer Affairs*, 41(1): 1–28.

Evaluation and Training Institute ETI (2008). JA finance park. Final report. Los Angeles: ETI.

Field, E., Jayachandran, S., and Pande, R. (2010). Do traditional institutions constrain female entrepreneurship? A field experiment on business training in india. *American Economic Review: Papers and Proceedings*, 100(2): 125–29.

Fort, M., Manaresi, F., and Trucchi, S. (2016). Adult financial literacy and households' financial assets: The role of bank information policies. *Economic Policy*, 31(88):743–782.

Flory, J. (2016) Formal Finance and Informal Safety Nets of the Poor: Evidence from a Savings Field Experiment, *mimeo*.

Garman, E.T., Kim, J., Kratzer, C.Y., Bruce, H.B., and Joo, S. (1999). Workplace financial education improves personal Financial Wellness. *Financial Counseling and Planning,* 10(1): 79–88.

Gaurav, S., Cole, S., and Tobacman, J. (2011). Marketing complex financial products in emerging markets: Evidence from rainfall insurance in India. *Journal of Marketing Research*, 48(SPL): S150–S162.

Gerrans, O. and Heaney, R. (2016). **T**he impact of undergraduate personal finance education on individual financial literacy, attitudes and intentions. *Accounting & Finance, forthcoming.*

Gibson, J., McKenzie, D., and Zia, B. (2014). The impact of financial literacy training for migrants. *World Bank Economic Review*, 28(1): 130–161.

Gill, A. and Bhattacharya, R. (2015). Integration of a financial literacy curriculum in a high school economics class: Implications of varying the input mix from an experiment. *Journal of Consumer Affairs*, 49 (2): 472–487.

Gine, X. and Mansuri, G. (2014). Money or ideas? A field experiment on constraints to entrepreneurship in rural Pakistan. *World Bank Policy Research Working Paper 6959*.

Gine, X., Karlan, D., and Ngatia, M. (2013). Social networks, financial literacy and index insurance. *World Bank, Washington, DC.*

Go, C.G., Varcoe, K., Eng, T., Pho, W., and Choi, L. (2012). Money savvy youth: Evaluating the effectiveness of financial education for fourth and fifth graders. *Federal Reserve Bank of San Francisco Working Paper 2012-02.*

Goda, G.S., Manchester, C.F., and Sojourner, A.J. (2014). What will my account really be worth? Experimental evidence on how retirement income projections affect saving. *Journal of Public Economics*, 119: 80–92.

Goldsmith, R.E. and Goldsmith, E.B. (2006). The effects of investment education on gender differences in financial knowledge. *Journal of Personal Finance*, 5(2): 55–69.

Grimes, P.W., Rogers, K.E., and Smith, R.C. (2010). High school economic education and access to financial services. *Journal of Consumer Affairs*, 44(2): 317–335.

Grinstein-Weiss, M., Guo, S., Reinertson, V., and Russel, B. (2015). Financial education and savings outcomes for low-income IDA participants: Does age make a difference? *Journal of Consumer Affairs*, 49(1): 156–185.

Han, C.-K., Grinstein-Weiss, M., and Sherraden, M. (2009). Assets beyond savings in individual development accounts. *Social Service Review*, 83(2): 221–244.

Hartarska, V. and Gonzalez-Vega, C. (2005). Credit counseling and mortgage termination by low-income households. *Journal of Real Estate Finance and Economics*, 30(3): 227–243.

Hartarska, V. and Gonzalez-Vega, C. (2006). Evidence on the effect of credit counseling on mortgage loan default by low-income households. *Journal of Housing Economics*, 15(1): 63–79.

Harter, C.L. and Harter, J.F. (2009). Assesing the effectiveness of Financial Fitness for life in eastern Kentucky. *Journal of Applied Economics and Policy*, 28: 20–33.

Harter, C.L. and Harter, J.F. (2010). Is financial literacy improved by participating in a stock market game? *Journal for Economic Educators*, 10(1): 21–32.

Haynes, D.C., Haynes, G., and Weinert, C. (2011). Outcomes of on-line financial education for chronically ill rural women. *Journal of Financial Counseling and Planning*, 22(1): 3–17.

Haynes-Bordas, R., Kiss, D., and Yilmazer, T. (2008). Effectiveness of financial education on financial management behavior and account usage: Evidence from a 'second chance' program. *Journal of Family and Economic Issues*, 29(3): 362–390.

Heinberg, A., Hung, A.A., Kapteyn, A., Lusardi, A., Samek, A.S., and Yoong, J. (2014). Five steps to planning success. Experimental evidence from U.S. households. *Oxford Review of Economic Policy,* 30(4): 697-724.

Hershey, D.A., Mowen, J.C., and Jacobs-Lawson, J.M. (2003). An experimental comparison of retirement planning intervention seminars. *Educational Gerontology*, 29(4): 339–359.

Hirad, A. and Zorn, P.M. (2001). A little knowledge is a good thing: Empirical evidence of the effectiveness of pre-purchase homeownership counseling. *Low-Income Homeownership Working Paper Series LIHO-01.4, JCHS, Harvard University.*

Hospido, L., Villanueva, E., and Zamarro, G. (2015). Finance for all: The impact of financial literacy training in compulsory secondary education in Spain. *Banco de España Working Paper 1502*, Madrid.

Jamison, J.C., Karlan, D, and Zinman, J. (2014). Financial education and acces to savings accounts: Complements or substitutes? Evidence from Ugandan youth clubs. *NBER Working Paper 20135.*

Kimball, M.S. and Shumway, T. (2010). Investor sophistication and the home bias, diversification, and employer stock puzzles. *Working Paper, University of Michigan.*

Krause, B. L., McCarthy, A. S., and Chapman, D. (2016). Fuelling financial literacy: estimating the impact of youth entrepreneurship training in Tanzania. *Journal of Development Effectiveness,* 8(2): 234–256.

Loke, V., Choi, L., and Libby, M. (2015). Increasing youth financial capability: An evaluation of the mypath savings initiative. *Journal of Consumer Affairs*, 49(1): 97–126.

Lusardi, A. (2002). Preparing for retirement: The importance of planning costs. *National Tax Association Proceedings–2002*: 148–154.

Lusardi, A. (2005). Financial education and the saving behavior of African American and Hispanic households, *Report*, *US Department of Labor, Employee Benefits Security Administration.*

Lusardi, A. and Mitchell, O.S. (2007). Financial literacy and retirement planning: New evidence from the Rand American life panel. *Michigan Retirement Research Center Research Paper No. WP 2007-157.*

Lusardi, A., Samek, A.S., Kapteyn, A., Glinert, L., Hung, A., and Heinberg, A. (2015). Visual tools and narratives: New ways to improve financial literacy. *Forthcoming in Journal of Pension Economics and Finance.*

Lührmann, M., Serra-Garcia, M., and Winter, J. (2015). Teaching teenagers in finance: Does it work? *Journal of Banking and Finance*, 54: 160–174.

Maki, D. (2004), Financial education and private pensions, in Gale, W., Shoven, J. and Warshowsky, M. (editors.), *Private Pensions and Public Policies*, Washington, DC: Brookings Institution Press: 126-139.☒

Mandell, L. (2006). Teaching young dogs old tricks:The effectiveness of financial literacy intervention in pre-high school grades. P*aper Presented at the Academy of Financial Services 2006 Annual Conference Salt Lake City.*

Mandell, L. (2009a). Starting younger: Evidence supporting the effectiveness of personal financial education for pre-high school students. *Working Paper, Aspen Institute and University of Washington.*

Mandell, L. (2009b). The impact of financial education in high school and college on financial literacy and subsequent financial decision making. *Paper presented at the American Economic Association Meetings, San Francisco, CA.*

Mandell, L. and Schmid Klein, L. (2009). The impact of financial literacy education on subsequent financial behavior. *Journal of Financial Counseling and Planning*, 20(1): 15–24.

Mills, G., Patterson, R. Orr, L., and DeMarco, D. (2004). *Evaluation of the american dream demonstration: Final evaluation report*, Cambridge, MA.

Muller, L. (2003). Does retirement education teach people to save pension distributions? *Social Security Bulletin,* 64 (4), 48–65.

Pang, M.F. (2010). Boosting financial literacy: benefits from learning study. *Instructional Science,* 38(6): 659–677.

Peng, T.-C.M., Bartholomae, S., Fox, J.J., and Cravener, G. (2007). The impact of personal finance education delivered in high school and college courses. *Journal of Family and Economic Issues*, 28(2): 265–284.

Quercia, R. and Spader, J. (2008). Does homeownership counseling affect the prepayment and default behavior of affordable mortgage borrowers? *Journal of Policy Analysis and Management*, 27(2): 304–325.

Reich, C.M. and Berman, J.S. (2015). Do financial literacy classes help? An experimental assessment in a low-income population. *Journal of Social Service Research*, 41(2): 193–203.

Reilly, B. D. (2016). Does mandatory financial education work? evaluating the impacts of mandatory secondary school standards on financial literacy. *Unpublished manuscript.*

Romagnoli, A. and Trifilidis, M. (2013). Does financial education at school work? Evidence from Italy. *Banco D'Italia Occasional Papers N. 155.*

Sanders, C.K., Weaver, T.L., and Schnabel, M. (2007) Economic education for battered women. An evaluation of outcomes. *Affilia: Journal of Women and Social Work*, 22(3): 240–254.

Sarr, L., Sadhu, S., and Fiala, N. (2012). Bringing the bank to the doorstep: Does financial education influence savings behavior among the poor? Evidence from a randomized financial literacy program in India. *Working Paper. Institute for Financial Management and Research, Centre for Micro Finance, German Instiute for Economic Research and the World Bank, Washington, D.C.*

Sayinzoga, A., Bulte, E. H., and Lensink, R. (2016). Financial literacy and financial behaviour: Experimental evidence from rural Rwanda. *Economic Journal*, 126(594): 1571–1599.

Schreiner, M., Sherraden, M., Clancy, M., Johnson, L., Curley, J., Grinstein-Weiss, M., Zhan, M., and Beverly, S. (2001). Savings and asset accumulation in individual development accounts. *Working paper, Center for Social Development, Washington University.*

Seshan, G. and Yang, D. (2014). Motivating migrants: A field experiment on financial decision-making in transnational households. *NBER Working Paper 19805.*

Skimmyhorn, W. (2016). Assessing financial education: Evidence from boot camp. *American Economic Journal: Economic Policy,* 8(2): 322–343.

Skimmyhorn, W. L., Davies, E. R., Mun, D., and Mitchell, B. (2016). Assessing financial education methods: Principles vs. rules-of-thumb approaches. *Journal of Economic Education*, 47(3): 193–210.

Song, C. (2012). Financial illiteracy and pension contributions: A field experiment on compound interest in China. *Unpublished Manuscript.*

Supanantaroek, S., Lensink, R., and Hansen, N. (2016). The impact of social and financial education on savings attitudes and behavior among primary school children in Uganda. *Evaluation Review, forthcoming.*

Tennyson, S. and Nguyen, C. (2001). State curriculum mandates and student knowledge of personal finance. *Journal of Consumer Affairs*, 35(2): 241–262.

Varcoe, K.P., Martin, A., Devitto, Z, and Go, C. (2005). Using a financial education curriculum for teens. *Journal of Financial Counseling and Planning*, 16(1): 63–71.

Walstad, W.B., Rebeck, K., and MacDonald, R.A. (2010). The effects of financial education on the financial knowledge of high school students. *Journal of Consumer Affairs*, 44(2): 336–357.

Wiener, R.L., Baron-Donovan, C., Gross, K., and Block-Lieb, S. (2005). Debtor education, financial literacy, and pending bankruptcy legislation. *Behavioral Sciences and the Law*, 23(3):

347–366.

Xiao, J.J., Serido, J., and Shim, S. (2012). Financial education, financial knowledge and risky credit behavior of college students. In: Lamdin, D.J. (editor) *Consumer Knowledge and Financial Decisions: Lifespan Perspectives*. New York: Springer: 113–128.

Yetter, E.A. and Suiter, M. (2015). Financial literacy in the community college classroom: A curriculum intervention study. *Federal Reserve Bank of St. Louis Working Paper 2015-001.*