NBER WORKING PAPER SERIES

FINANCIAL EDUCATION AFFECTS FINANCIAL KNOWLEDGE AND DOWNSTREAM BEHAVIORS

Tim Kaiser
Annamaria Lusardi
Lukas Menkhoff
Carly J. Urban

Working Paper 27057
http://www.nber.org/papers/w27057

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2020

We thank participants of the 5th Cherry Blossom Financial Education Institute in Washington, D.C., and Michael Collins, Andrea Hasler, Rachael Meager, Olivia Mitchell, and Pierre-Carl Michaud for many helpful comments. We thank Shawn Cole, Daniel Fernandes, Xavier Giné, John Lynch, Richard Netemeyer, and Bilal Zia for providing details about their studies. Financial support by DFG through CRC TRR 190 is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Financial Education Affects Financial Knowledge and Downstream Behaviors
Tim Kaiser, Annamaria Lusardi, Lukas Menkhoff, and Carly J. Urban
NBER Working Paper No. 27057
April 2020
JEL No. D14,G53,I21

ABSTRACT

We study the rapidly growing literature on the causal effects of financial education programs in a meta-analysis of 76 randomized experiments with a total sample size of over 160,000 individuals. The evidence shows that financial education programs have, on average, positive causal treatment effects on financial knowledge and downstream financial behaviors. Treatment effects are economically meaningful in size, similar to those realized by educational interventions in other domains, and are at least three times as large as the average effect documented in earlier work. These results are robust to the method used, restricting the sample to papers published in top economics journals, including only studies with adequate power, and accounting for publication selection bias in the literature. We conclude with a discussion of the cost-effectiveness of financial education interventions.

Tim Kaiser
Department of Economics
University of Koblenz-Landau
Landau 76829
Germany
and DIW Berlin
kaiser@uni-landau.de

Annamaria Lusardi
The George Washington University
School of Business
2201 G Street, NW
Duques Hall, Suite 450E
Washington, DC 20052
and NBER
alusardi@gwu.edu

Lukas Menkhoff
Humboldt-University of Berlin
Germany
and DIW Berlin
lmenkhoff@diw.de

Carly J. Urban
Department of Agricultural Economics
and Economics
Montana State University
P.O. Box 172920
Bozeman, MT 59717
and Institute for Labor Studies (IZA)
carly.urban@montana.edu

A data appendix is available at http://www.nber.org/data-appendix/w27057

1 Introduction

The economic importance of financial literacy is documented in a large and growing empirical literature (Hastings et al. 2013; Lusardi and Mitchell 2014; Lusardi et al. 2017; Lührmann et al. 2018). Consequently, the implementation of national strategies promoting financial literacy and the design of financial education initiatives and school mandates have become a high priority around the world. Many of the largest economies, including most OECD member countries, as well as India and China, have implemented programs enhancing financial education in order to promote financial inclusion and financial stability (OECD 2015). Together, these financial education programs seek to reach more than five billion people in sixty countries, and the number of countries joining this effort continues to grow.

Despite the many initiatives to foster financial literacy, the effectiveness of financial education is debated in quite fundamental ways. Much of the debate stems from the fact that the limited number of early rigorous experimental impact evaluations sometimes showed muted effects, and these early findings have contributed to the perception of mixed evidence on the effectiveness of financial education (see, for example, Fernandes et al. 2014). However, empirical studies on financial education have grown rapidly in the past few years. To account for the large increase in research in this field, we take stock of the recent empirical evidence documented in randomized experiments and provide an updated and more sophisticated analysis of the existing work.

Our main finding is clear-cut: financial education in 76 randomized experiments with a total sample size of more than 160,000 individuals has positive causal treatment effects on financial knowledge and financial behaviors. The treatment effects on financial knowledge are similar in magnitude to the average effect sizes realized by educational interventions in other domains, such as math and reading (see Hill et al. 2008; Cheung and Slavin 2016; Fryer 2016; Kraft 2019). The effect sizes of financial education on financial behaviors are comparable to those realized in behavior-change interventions in the health domain (e.g., Rooney and Murray 1996; Portnoy et al. 2008; Noar et al. 2017) or behavior-change interventions aimed at fostering energy conserving behavior (e.g., Karlin et al. 2015).

Specifically, the estimated (weighted average) treatment effect is *at least three times as large* as the weighted average effect documented in Fernandes et al. (2014), which examined 13 Randomized Controlled Trials (RCTs). The analysis from our more sophisticated meta-analysis, which accounts for the possibility of cross study heterogeneity, results in an estimated effect of financial education interventions that is more than *five times as large* as the effect reported in Fernandes et al. (2014).

Additionally, we calculate the effect sizes resulting from these interventions and show that they are of economic significance. Our results are robust, irrespective of the model used, when restricting the sample to only those RCTs that have been published in top economics journals, when restricting the sample to only those studies with adequate power to identify small treatment effects, and when employing an econometric method to account for the possibility of publication selection bias favoring the publication of statistically significant results.

In contrast to earlier studies, we do not find differences in treatment effects for low-income individuals and the general population. We also do not find strong evidence to support a rapid decay in the realized treatment effects, though we do not find support for the sustainability of long-run effects either.

For completeness and to asses the external validity of the findings, we also discuss the findings from recent evaluations of financial education mandates and school financial education programs operated at scale.

With this work, we make four main contributions. First, we provide the most comprehensive analysis of the burgeoning work on financial education by using the most rigorous studies: randomized control trials. Second, we focus on a critical feature of empirical analyses on micro data: the heterogeneity in the programs and the many differences that normally one finds in the programs; for example, differences in target groups, quality and the intensity of interventions. Third, we discuss the magnitudes of the effects in terms of economic significance and consider the per participant costs of programs. Fourth, we provide a thorough discussion of topics raised in previous work, i.e., how to assess the impact of financial education and whether education decays with time. We believe that this work can provide useful guidance for those evaluating future financial education programs.

The paper has seven sections: section 2 serves as a primer on statistical meta-analyses; section 3 describes our method; section 4 presents descriptive statistics of our data; section 5 presents the results of our analyses; section 6 discusses the economic significance of our effect sizes and the cost-effectiveness associated with these effects; section 7 concludes.

2 Background

As the amount of evidence from rigorous empirical studies in a given field grows over time, there is an increased need to synthesize and integrate the existing findings to reach a consistent conclusion. Traditionally, economists have relied on narrative reviews, where experts on a given literature select and discuss the most relevant findings. The advantage of such an approach is that the experts are expected to have a good understanding of the existing studies and can add value by summarizing, interpreting, and linking together the most convincing (internally valid) studies in a narrative review. Examples of widely cited narrative reviews in the financial education literature are Fox et al. (2005), Collins and O'Rourke (2010), Xu and Zia (2012), Hastings et al. (2013), and Lusardi and Mitchell (2014).

As empirical literatures grow larger, however, narrative literature reviews can become difficult, since it is hard to describe a large number of empirical estimates and discuss all of the possible sources of heterogeneity in reported findings. Meta-analyses have thus become more common in economics when aggregating findings from many studies. Some examples of recent meta-analyses in economics include Meager (2019), which studies microcredit expansions, and Beuermann and Jackson (2018), which examines the effect of going to parent-preferred schools. Meta-analyses can serve as a complement to narrative reviews when there is a sufficiently large number of well-identified studies on the same empirical research question. A meta-analysis—a systematic, quantitative literature review—is well suited to obtain an estimate of the average effects of a given program and to study the heterogeneity in reported findings (Stanley 2001).

As noted earlier, Fernandes et al. (2014) was the first meta-analysis performed in the field of financial education. We differ from this initial and well-cited study in three major ways. First, we update the dataset to incorporate the many papers that have been written since the meta-analysis by Fernandes et al. was published. As Figure 1 shows, the field grew exponentially after 2014, so previous reviews cover only a small part of the work that currently exists. Second, we attempt to replicate the findings in Fernandes et al. (2014), and we provide estimates more common in meta-analysis literature, which account for heterogeneity in effect sizes across studies. This takes into consideration, for example, the intensity of the program. Third, we have chosen to focus solely on what are considered the most rigorous sources of evidence, i.e., randomized experiments. RCTs provide more consistent internal validity than observational and quasi-experimental studies, especially since there are no universally accepted instruments for financial literacy, and one can debate whether existing non-randomized trials have made use of convincing empirical strategies addressing endogeneity of selection into treatment. Judging the quality of quasi-experimental studies and determining which to include or exclude from the meta-analysis gives researchers an additional degree of freedom that we wish to remove. Importantly, the number of RCTs has grown from just 13 in the Fernandes et al. (2014) review to 76 as of 2019. In those 13 studies, the authors found the weakest effects of financial education interventions reviewed in their work. Fernandes et al. (2014) assert that these studies provide the strongest evidence against financial education.

< Figure 1 about here >

In addition to Fernandes et al. (2014), there have been three follow-up meta-analyses on financial education programs: Miller et al. (2015); Kaiser and Menkhoff (2017); and Kaiser and Menkhoff (2019). These meta-analyses present a more nuanced view of financial education interventions than the original paper by Fernandes et al (2014) by including additional studies and accounting for differences in program design and outcomes studied. This study will build upon those, but it expands the contribution by focusing solely on RCTs, including additional years of data, deepening the methodological discussion (including new robustness checks), providing a thorough discussion of economic significance, and incorporating information on program costs. By contrast, Miller et al. (2015) focus on less than 20 studies and put emphasis on examining impact differences across outcomes. Kaiser and Menkhoff (2017) concentrate on the determinants of effective financial education interventions, while Kaiser and Menkhoff (2019) focus on financial education interventions in schools.

3 Methods

This section describes our inclusion criteria for the papers on financial education (Section 3.1), the details we use in constructing our database of effect sizes (Section 3.2), and the specifics of the empirical model we employ (Section 3.3).

3.1 Inclusion criteria

In order to draw general conclusions about a given literature, one has to conduct a systematic search of the literature and apply inclusion criteria that are defined ex-ante. We conducted a search of all relevant databases for journal articles and working papers (see

Appendix A for the list of the studies we considered and a summary of the data we extracted from those studies), and apply three inclusion criteria to the universe of records return in this set. Criteria of inclusion: (i) Studies reporting the causal effects of educational interventions designed to strengthen the participants' financial literacy and/or leading to behavior change in the area of personal finance; (ii) studies using random assignment into treatment and control conditions; (iii) studies providing a quantitative assessment of intervention impact that allows researchers to code an effect size estimate and its standard error. Where necessary information is partially missing, we consulted additional online resources related to the article or contacted the authors of the studies. We only consider the main results discussed in the text, and we do not code redundant effect sizes (e.g., effect sizes arising from other specifications of a given statistical model in the robustness section). Table A1 provides a list of all the studies considered in our analysis.

3.2 Constructing the database

Our analysis aggregates treatment effects of financial education interventions into two main categories. First, we code the effect of financial education on *financial literacy* (i.e., a measure of performance on a financial knowledge test) since improvement in knowledge is usually the primary goal of financial education programs (Hastings et al. 2013; Lusardi and Mitchell 2014) and is expected to be one of the channels via which financial behavior is influenced. We do not include self-assessments of changes in financial knowledge as an outcome.

Second, we code the effect of financial education on *financial behaviors.* These behaviors can be further disaggregated into the following categories: Borrowing, (retirement) saving, budgeting and planning, insurance, and remittances. It is useful to know, for example, which behavior is more easily impacted by financial education. Table A3 provides an overview of the categories and definitions of outcome types.

We code all available effect sizes per study on financial knowledge and behavioral outcomes. We include multiple estimates per study if multiple outcomes, survey-rounds, or treatments are reported. We only extract main treatment effects reported in the papers. Thus, we do not consider 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, etc.), with the treatment indicators. We aim to only consider intention-to-treat effects (ITT), unless these are not reported. If only local average treatment effects (LATE) or the treatment effect on the treated (TOT) are reported, we included these in our analysis and check for statistical differences, as described in Appendix B.[1]

This process leads to the inclusion of 76 independent randomized experiments described further in Section 4.

3.3 Empirical model

A major challenge in every meta-analysis lies in the heterogeneity of the underlying primary studies and how to account for it. In the financial education literature, heterogeneity arises from several sources; in our sample, randomized experiments on financial education programs have been conducted in 32 countries with varying target groups (see Table A1 in Appendix A). Moreover, the underlying educational interventions are very diverse, ranging from provision of an informational brochure to offering high-intensity classroom instruction; outcomes are also measured at different points in time and with different types of data. Accommodating this heterogeneity is important in order to draw general conclusions about the findings.

[1] We also show results for the sample of studies reporting the ITT in Appendix B.

When there is such heterogeneity in the studies under consideration, meta-analyses require certain assumptions about the sources of variance in the observed treatment effect estimates. Consider a set of j randomized experiments, each of them reporting an estimate of a causal (intention to treat) treatment effect relative to a control group.[2] Assuming no heterogeneity in true effects implies that the observed estimates of a treatment effect are sampled from a distribution with a single true effect β_0 and variance σ^2, as in the following meta-analysis model:

$$y_j = \beta_0 + \epsilon_j \tag{1}$$

where y_j is an estimate of a treatment effect in the jth study, β_0 defines the common true effect, and ϵ_j is the study level residual with $\epsilon_j \sim N(0, \sigma_j^2)$. Thus, the estimate of the common true effect is given by estimating the above model with weighted least squares using inverse variance weights ($w_j = \frac{1}{\sigma_j^2}$). While this may be a reasonable assumption for some empirical literatures, such as medical trials with identical treatment, dosage, and procedures for measuring outcomes, this is clearly not a reasonable assumption in the context of educational interventions, which tend to be quite diverse.

A more reasonable approach in an educational setting would be to assume heterogeneity between studies, hence assuming a distribution of possible true effects, allowing true effects to vary across studies with identical within-study measurement error. The weighted average effect

[2] Because each study j may report its treatment effect estimate in a different unit (i.e., a different currency or on different scales), we convert each estimate to a (bias corrected) standardized mean difference (Hedges' g), such that the treatment effect estimate y_j is standardized as $g_j = \frac{\mathrm{M}_T - \mathrm{M}_\mathrm{C}}{\mathrm{SD_p}}$ with $SD_p = \sqrt{\frac{(\mathrm{n_T}-1)\, SD_T{}^2 + (\mathrm{n_C}-1)\, SD_C{}^2}{n_T{}^2 + n_C{}^2 - 2}}$, i.e., the mean difference in outcomes between treatment (M_T) and control ($\mathrm{M_C}$) as a proportion of the pooled standard deviation (SD_p) of the dependent variable. $\mathrm{n_T}$ 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, the standard error of each standardized mean difference is defined as: $SE_{g_j} = \sqrt{\frac{\mathrm{n_T}+\mathrm{n_C}}{\mathrm{n_T}\mathrm{n_C}} + \frac{g_j^2}{2(\mathrm{n_T}+\mathrm{n_C})}}$.

then does not represent a single true effect but instead the mean of the distribution of true effects. Thus, the model can be written as:

$$y_j = \beta_0 + \upsilon_j + \epsilon_j \tag{2}$$

with $\upsilon_j \sim N(0, \tau^2)$ and $\epsilon_j \sim N(0, \sigma_j^2)$. τ^2 is the between-study variance in true effects that is unknown and has to be estimated from the data,[3] and σ_j is the within-study standard error of the treatment effect estimate y_j that is observed for each study j. Subsequently, weighted least squares is used to estimate β_0 with inverse variance weights defined as $w_j = (\tau^2 + \sigma_j^2)^{-1}$. Thus, instead of estimating one common effect, the goal is to estimate the mean of the distribution of true effects.

While the illustration so far has considered cases in which each study contributes one independent treatment-effect estimate, this is generally not the case in the financial education literature. Instead, studies may report treatment effect estimates from multiple treatments and a common control group within studies, at multiple time-points and for multiple outcomes. Therefore, we extend the model above to incorporate multiple (and potentially correlated) treatment effect estimates within studies:

$$y_{ij} = \beta_0 + \upsilon_j + \epsilon_{ij} \tag{3}$$

y_{ij} is the ith treatment effect estimate within each study j. β_0 is the mean of the distribution of true effects, $\upsilon_j$is the study-level random effect with $\upsilon_j \sim N(0, \tau^2)$, τ^2 is the between study variance in true effects, and $\epsilon_{ij} \sim N(0, \sigma_{ij}^2)$ is the residual of the ith treatment effect estimate within each study j. This model allows between-study heterogeneity in true effects but assumes that treatment effect estimates within studies relate to the same study-specific true effect. This

[3] There are several possible algorithms to estimate the between-study variance τ^2. Our approach uses the method of moments estimator (see Harbord and Higgins 2008), but iterative approaches, such as (restricted) maximum likelihood or empirical Bayes estimation, are also frequently used in meta-analyses.

means the common within-study correlation of treatment effect estimates is induced by random sampling error.

While the estimator proposed in Hedges et al. (2010) does not require an exact model of the within-study dependencies in true effects, Tanner-Smith and Tipton (2014) and Tanner-Smith et al. (2016) suggest that the following inverse variance weights (w_{ij}) are approximately efficient in case of a correlated effects model:

$w_{ij} = \left\{\left(\tau^2 + \frac{1}{k_j}\sum_{k_j=1}^{k_i} \sigma_{ij}^2\right)\left[1 + \left(k_j - 1\right)\rho\right]\right\}^{-1}$, where τ^2 is the estimated between-study variance in true effects, $(\frac{1}{k_j}\sum_{k_j=1}^{k_i} \sigma_{ij}^2)$ is the arithmetic mean of the within-study sampling variances (σ_{ij}^2) with k_j being the number of i effect size estimates within each study j, and ρ is the assumed common within-study correlation of treatment effect estimates.

We estimate the model with these weights and choose ρ = 0.8 as the default within-study correlation of estimates (see Tanner-Smith and Tipton 2014). However, sensitivity analyses of such an assumption are easily implemented, and we show results for ρ = [0, 0.9] in increments of 0.1 in Appendix B.

Our method addresses several shortcomings of the analysis presented in Fernandes et al. (2014). First, we are able to formally investigate the importance of modeling between-study heterogeneity in treatment effects and to compare the results to a model with the common-effect assumption used in Fernandes et al. (2014). This is important because, as mentioned before, financial education programs can be very different from each other. Second, we make use of the all of the statistical information reported in primary studies, since the method used in this paper is able to accommodate multiple estimates within studies, and thus is not dependent on creating highly aggregated measures, such as the within-study average effect sizes reported in Fernandes et al. (2014). To probe the robustness of our results, we estimate five alternative models (see Appendix B), including a correction for potential publication selection bias and a consideration of the power of the underlying primary studies. We are also careful to replicate the methods of Fernandes et al. (2014), as reported in Appendix D.

4 Data

To arrive at an unbiased estimate of the mean of the distribution of true effects of financial education programs, we collect a complete list of randomized experiments in the financial education literature. We build on an existing database and update it using the search strategy described earlier, which is also used in Kaiser and Menkhoff (2017). We augment the earlier dataset used in previous work with published randomized experiments on financial education through January 2019 (end of collection period for this paper).[4] Appendix A contains a detailed description of the papers included in our meta-analysis and the types of outcomes coded. Applying our inclusion criteria, we arrive at a dataset of as many as 68 papers reporting the effects of 76 independent-sample experiments. This is a much bigger sample of RCTs than any previous meta-analyses.

An important part of our meta-analysis is the inclusion of many recent papers in our dataset, which enables us to provide a comprehensive and updated review of the large and rapidly growing amount of research done on this topic. The review by Fernandes et al. (2014) is the first paper in the literature, and it covers only 13 RCTs from which they code 15 observations. The meta-analysis in Miller et al. (2015) covers a total of seven RCTs. Of our 76 independent-sample experiments, one-third have not been included in the most recent meta-analysis, by Kaiser and Menkhoff, (2017).[5] Thus, we expand greatly on those previous studies. Table C1 in Appendix C contains a comparison of our dataset of RCTs to these earlier accounts of the literature.

[4] This paper has gone through revisions and the end of the collection period refers to when we started extracting and analyzing the data.
[5] We are also careful to update all of the papers to the latest version and include, for example, the estimates in the published version of the papers.

From our sample of 76 independent randomized experiments, we extract a total of 673 estimates of the effects of the program (the treatment effects). Out of these, 64 studies report a total of 458 treatment effects on financial behaviors (see Table A4 in Appendix A). Thus, we are able to work on a large number of estimates. The studies vary in their choice of dependent variables, ranging from a number of financial behaviors to financial knowledge. To illustrate some simple differences in studies, we note that 23 studies report 115 treatment effect estimates on *credit behaviors*, and 23 studies report 55 treatment effect estimates on *budgeting behavior*. The largest number of estimates is on *saving behavior*, with 54 studies reporting a total of 253 treatment effect estimates. Six studies report 18 treatment effect estimates on *insurance behavior*, and six studies report 17 estimates on remittance behavior. Fifty studies report 215 treatment effect estimates on *financial knowledge* and 38 studies report treatment effects on both knowledge and behaviors. We have a sizeable number of estimates for each outcome.

We start our analysis by showing that the descriptive statistics alone suggest that financial education is, on average, effective in improving both knowledge and behavior.

< Table 1 about here >

The average effect size across all types of outcomes, reported in Table 1, is 0.123 standard deviation (SD) units (SD=0.183), and the median effect size is 0.098 SD units.[6] The minimum effect size is -0.413, and the maximum effect size is 1.374. The average standard error of the treatment effect is 0.085 (SD=0.049) and the median standard error is at 0.072.[7]

We first note that there is substantial variation in instruction time in the programs, where the average estimate is associated with a mean of 11.71 hours of instruction (SD=16.27), and the median is associated with 7 hours of instruction (Table 1). Treatment effects are estimated 30.4 weeks (7 months) after treatment, on average, with a standard deviation of 31.65 weeks

[6] Note that all effect sizes are scaled such that desirable outcomes have a positive sign (i.e., we are coding a negative coefficient on “loan default” as a positive treatment effect (i.e., reduction in loan default) and vice versa.
[7] The average sample size across the 76 randomized experiments is 2,136 and the median sample size is 840.

(7.3 months). The median study does not focus on immediate effects: the median time passed between financial education treatment and measurement of outcomes is 25.8 weeks (5.9 months). This is useful information for assessing the impact of programs, in particular if one hypothesizes a decay of effectiveness with time, as emphasized by Fernandes et al. (2014). Further, we note that nearly three quarters (72.4 percent) of the treatment effect estimates target low-income individuals (income below the median), and 60.8 percent of the estimates are from programs studied in developing economies; 30.8 percent of all estimates reported in randomized experiments appear in top economics journals, which reflects the high quality of this sample of studies. The average age across all reported estimates is 33.5 years, where 7.5 percent of estimates are focused on children (<14 years old), 20 percent are focused on youth (14-25 years old), and 72.4 percent are focused on adults (>25 years old).

When assessing the effectiveness of financial education, interventions may not necessarily lead to changes in behavior if people have resource constraints or are in the early part of the life cycle, as highlighted in Lusardi et al. (2017). In some cases, people may already be acting optimally and in other cases, even after exposure to financial education, it may be optimal to not change behavior. Determining which behaviors should optimally change requires a theoretical framework sometimes lacking in this literature.

5 Results

We present the results in three steps. Section 5.1 shows the main results of our meta-analysis of the universe of randomized experiments (up to 2019) and compares the results to the first meta-analysis of the literature by Fernandes et al. (2014). Section 5.2 summarizes the results of comprehensive robustness exercises that are reported in full in Appendix B. Section 5.3 examines our main effects further by discussing the results by outcomes, such as financial knowledge and a variety of financial behaviors. Section 5.4 presents our main results once we disaggregate the data into various sub-samples of interest.

5.1 A meta-analysis of randomized experiments

We describe our findings by first plotting the universe of 673 raw effects extracted from the 76 studies against their inverse standard error (precision) in Figure 2. We disaggregate the data and distinguish between estimated treatment effects on *financial behaviors* (n=458) and *financial knowledge* (n=215).[8] The unweighted average effect on financial behaviors is 0.0898 SD units, and the unweighted average effect on financial knowledge is 0.187 SD units. With this simple analysis of the raw data, we find that financial education improves both financial knowledge and behaviors.

A visual inspection of the plot in Figure 2 shows that both samples of effect sizes resemble a roughly symmetric funnel until effect sizes of 0.5 SD units and above. We investigate the possibility of publication selection bias[9] in the financial education literature in Appendix B (see Figure B1 and Table B1) and find that accounting for this potential publication bias does not qualitatively change the result of positive average effects of financial education.

< Figure 2 about here >

Next, we provide a comparison of the data in our study with the results in Fernandes et al. (2014). Specifically, we estimate the weighted average effect on financial behaviors using 'Robust Variance Estimation in Meta-Regression with Dependent Effect Size Estimates' (RVE) under the common true effect assumption[10] made in Fernandes et al. (2014) and compare our

[8] We refer to n as the number of estimates and not the number of participants in the studies.
[9] Publication selection bias refers to the potential behavior of researchers to be more likely to report and journal editors being more likely to publish statistically significant results.
[10] Thus, we assume $\tau^2 = 0$, i.e., the weights are defined as $w_{ij} = \left\{\left(\frac{1}{k_j}\sum_{k_j=1}^{k_i}\sigma_{ij}^2\right)\left[1+\left(k_j-1\right)\rho\right]\right\}^{-1}$. Note, that Fernandes et al. (2014) use only one observation per study by creating within-study average effect sizes, i.e., the weights in their study are defined as $w_j = \frac{1}{\sigma_j^2}$. We show results with this approach in Table B3 of Appendix B.

result in the larger sample of 64 RCTs to their earlier result based on 15 observations from 13 RCTs.[11] These results are reported in Figure 3.

< Figure 3 about here >

A few important clarifications are in order: Fernandes et al. (2014)'s estimate and standard error in Figure 3 is from the analysis of 15 observations of RCTs in their paper, not from our analysis of their data. We were not able to exactly replicate this result, and in the process, we uncovered four data errors in the direct coding and classification of RCT effect sizes. In Appendix D, we describe our attempt to replicate the original result by Fernandes et al. (2014) and thoroughly document each coding discrepancy.

Taking their estimates at face value, Figure 3 shows that simply updating the dataset to incorporate the burgeoning recent work increases the effect by more than three times. Compared to the estimate reported in Fernandes et al. (2014) of 0.018 SD units (with a 95% confidence interval (CI_{95}) from -0.004 to 0.022), the weighted average effect in this larger sample of recent RCTs is about 3.6 times higher. The new estimate of the effect size, even with the identical assumption of a common true effect, clearly rules out a null effect of financial education (0.065 SD units with CI_{95} from 0.043 to 0.089). Thus, one of the main findings of Fernandes et al. (2014) is not confirmed in this larger sample of RCTs.

Because the common true effect assumption is potentially problematic in the context of heterogeneous financial education interventions, we estimate the mean of a distribution of true effects using the model specified in equation 3. In addition to the mentioned theoretical reasons

[11] We convert the correlations used as an effect size metric by Fernandes et al. (2014), (*r)* to a standardized mean difference (Cohens' *d*) $d = \frac{2r}{\sqrt{1-r^2}}$ and we convert the standard error using $SE_d = \sqrt{\frac{4SE_r^2}{(1-r^2)^3}}$ (cf. Lipsey and Wilson 2001). This is true under the assumption that the outcome measures in each group are continuous and normally distributed and that the treatment variable is a binary variable indicating treatment and control groups, i.e., a valid assumption in the context of RCTs. To arrive at the "bias corrected standardized mean difference" (Hedges' g) one may apply the following bias correction factor ex post $g = d\,\left(1 - \frac{3}{4(n_1+n_2-2)-1}\right)$ (cf. Borenstein et al. 2009) but these metrics are near identical in the context of the financial education literature where the average sample size is 2,136 and the median sample size is 840.

to assume a distribution of true effects rather than a single true effect, we note that formal tests of heterogeneity show that at least 86.4 percent of the observed between-study variance can be attributed to heterogeneity in true effects and only 13.6 percent of the observed variance would have been expected to occur by within-study sampling error alone (see Table B3 in Appendix B).[12]

Figure 3 shows the result of the random-effects RVE model. In our view, this estimated mean of the distribution of financial education treatment effects is the most appropriate aggregate effect size to consider; the estimate results in a mean of 0.1003 SD units [CI_{95} from 0.071 to 0.129], and thus, is significantly different from the estimate using the common true effect assumption. The effect of financial education is now approximately 5.5 times larger than the estimate reported in Fernandes et al. (2014). This effect is very similar in magnitude to statistical effect sizes reported in meta-analyses of behavior-change interventions in other domains such as health (e.g., Rooney and Murray 1996; Portnoy et al. 2008; Noar et al. 2007) or energy conservation behavior (e.g., Karlin et al. 2015).

To summarize, evidence that incorporates the updated set of papers shows that financial education is effective, on average. Hence, we do not confirm the estimates from early studies, which are based on a small number of interventions.

5.2 Model sensitivity

We probe the robustness of our findings about the average effect of financial education programs with various sensitivity checks that are reported in full in Appendix B. These tests include (i) estimating three alternative meta-analyses including models with a common-effect assumption, (ii) investigating and correcting for potential publication selection bias, (iii) restricting the sample to only those studies with adequate power to identify small treatment

[12] A Cochran's Q-test of homogeneity (with one synthetic effect size per study) results in a Q-statistic of 464.71 ($p<0.000$).

effects, (iv) choosing different assumed within-study correlations of treatment effect estimates for the random-effects RVE approach, and (v) creating one synthetic effect size per study (inverse-variance weighted within-study average) and estimating both fixed-effect and random-effects models with one observation per study. All of these robustness checks confirm the main conclusions of our paper.[13]

5.3 Outcome domains

In addition to the effects on financial behaviors aggregated above (Figure 3), i.e., all behaviors, we also include estimates on financial knowledge (Figure 4). Treatment effects on *financial knowledge* are larger than the effect sizes on *financial behaviors*.

< Figure 4 about here >

Specifically, we find that the mean of the distribution of true effects in our sample is estimated to be 0.204 [CI_{95} from 0.152 to 0.255]. Hence, here as well, we cannot confirm the finding by Fernandes et al. (2014) based on 12 papers (average effect of about 0.133 SD units).[14] Instead, our average effect on financial knowledge is very similar to the average effects of educational interventions in math or reading (see Hill et al. 2008; Cheung and Slavin 2016; Fryer 2016; Kraft 2019).

Effect sizes on *financial behaviors* are mostly not statistically different from each other, suggesting the adequacy of pooling across these outcomes. However, additional analyses shown in Table B2 in Appendix B suggest that the results on *saving behavior* and *budgeting behavior* are the most robust, while the effects on other categories of financial behaviors are less certain due to either fewer studies including these outcomes (*insurance* and *remittances)* or high heterogeneity in the estimated treatment effects (*credit behaviors)*. This result is

[13] We also check the robustness of results when excluding any papers of the authors of this meta-analysis.
[14] See Fernandes et al. (2014), p. 1867: "In 12 papers reporting effects of interventions on both measured literacy (knowledge) and some downstream financial behavior, the interventions explained only 0.44% of the variance in financial knowledge," i.e., $\sqrt{r^2} = 0.066$ or d=0.133.

generally in line with earlier accounts of the literature, such as Fernandes et al. (2014), Miller et al. (2015), and Kaiser and Menkhoff (2017), and extends to the larger set of RCTs.

5.4 Subgroup analyses

In order to better understand the sources of heterogeneity in this literature, we further disaggregate our data into various subgroups and investigate the mean effect of financial education interventions.

5.4.1 Sample population

We disaggregate the sample of RCTs by characteristics of the sample population. First, we split the sample by country-level income, distinguishing between high income economies and developing economies, to account for differences in resources.[15] We find that the treatment effects of interventions in developing economies on financial behaviors are about 9.56 percent smaller than those in richer countries; however this difference is not statistically significant (see Panel A(a) of Table 2). Previous meta-analyses have found slightly smaller effect sizes for interventions in developing economies when controlling for additional features of the programs, such as intensity (cf. Kaiser and Menkhoff 2017). Treatment effects on financial knowledge are about 46 percent smaller in developing economies than in high income economies (see Panel B(a) of Table 2); this difference is statistically significant, and this is also in line with earlier evidence based on a smaller sample of RCTs (cf. Kaiser and Menkhoff 2017).

< Table 2 about here >

We next look at differences between low-income individuals and people with average or above average individual income (relative to the average within-country income). While

[15] Country groups are based on the World Bank Atlas method and refer to 2015 data on Gross National Income (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.

interventions with low-income individuals show smaller treatment effects, on average, which is in line with earlier accounts of the literature (Fernandes et al. 2014; Kaiser and Menkhoff 2017), we—in contrast to these earlier studies—do not find any significant differences between these two samples (see Panel A(b) and Panel B(b)); this indicates that recent RCTs added to the sample show smaller differences in treatment effects between groups than those interventions studied in the earlier literature.

Additionally, we disaggregate our sample by the age of the participants (see Panel A(c) and Panel B(c) of Table 2). Treatment effects on financial behaviors are smallest for children (below age 14) (0.064 SD units) relative to youth (ages 14 to 25) (0.1203 SD units) and adults (above age 25) (0.1068 SD units), while the latter difference is only marginally significant. Treatment effects on financial knowledge, on the other hand, are estimated to be largest among children (0.2763 SD units) relative to youth (0.1859 SD units) and adults (0.2001 SD units). These differences, however, are not statistically significant due to large uncertainty around the estimate for children, which is based on 15 observations in seven studies (CI_{95} from 0.0076 to 0.545).

5.4.2 Journal quality

To address possible concerns regarding the internal validity and general rigor of the included experiments and to focus on what editors and reviewers have judged to be the highest quality evidence, we restrict the sample to studies published in top general interest or top field economics journals only.[16] We compare the estimated treatment effects on financial behaviors of the 15 studies published in these journals to the estimated treatment effects of the other 49 studies published in other journals or as working papers. While treatment effects are estimated

[16] These journals are: (1) *Quarterly Journal of Economics*, (2) *Journal of Political Economy*, (3) *American Economic Journal: Applied Economics*, (4) *American Economic Journal: Economic Policy*, (5) *Journal of the European Economic Association*, (6) *Economic Journal*, (7) *Journal of Finance*, (8) *Review of Financial Studies*, (9) *Management Science*, (10) *Journal of Development Economics*. There were no publications in other top journals, such as the *American Economic Review, Econometrica, and the Review of Economic Studies*.

to be slightly smaller in these types of publications, there are no statistically significant differences between these types of publications (see Panel A(d) and Panel B(d) of Table 2). The same is true for effect sizes on financial knowledge where eight experiments published in top general interest or top field economics journals report smaller, albeit not statistically different, effect sizes than 42 experiments published in other journals or as working papers.

5.4.3 Time horizon

Finally, we tackle the important topic of potential decay of effectiveness of financial education over time. We disaggregate the sample of treatment effects within studies, considering the time span between financial education treatment and measurement of outcomes (see Panel A(e) and Panel B(e) of Table 2). We start by looking at treatment effect estimates that measure outcomes in the very short run (i.e., a time span of less than six months). The average effect of financial education on financial behaviors within this sample of 34 RCTs (180 effect sizes) is 0.0991. Looking at treatment effects on financial behaviors that are measured at a time span of six months or more (28 experiments and 260 estimates), we find that the estimates reduced to 0.071 SD units [CI_{95} from 0.0425 to 0.0995], which is a marginally significant difference relative to the set of studies with the shorter time horizon.

We next restrict the sample further to 18 studies that measure treatment effects on financial behaviors after at least one year. The estimate is statistically not different to the studies with shorter time horizon after treatment (0.0878 SD units). Restricting the sample to even longer time spans, i.e., ten RCTs that measure effects on financial behaviors at least 1.5 years after treatment or longer, results in an estimated average of 0.0653 SD units. These effects are slightly reduced but are still not statistically different from the other estimates. Restricting the set of RCTs further to those seven studies that measure treatment effects on financial behaviors at least two years after treatment or longer, results in an estimate of 0.0574 SD units, which is again not statistically different from the other estimates and does not include the possibility of zero effects (within the limits of the 95% CI). Overall, there is some decay in effectiveness when measurement is delayed by six months or more; however, beyond this threshold we do not observe any further significant decline.

Regarding the decay in financial knowledge, we find significantly larger effects (0.2305 SD units) in 36 RCTs measuring effects on financial knowledge in the very short run (i.e., at a time span shorter than six months) relative to those with time horizons above six months (0.1408 SD units), but no statistically significant differences at longer time horizons (more than 6 months or more than 12 months). However, only five studies measure treatment effects on financial knowledge considering time horizons between 12 and 18 months, and no longer-term studies exist in our sample.

Overall, these examinations of the possible decay in outcomes highlighted by Fernandes et al. (2014) do not find conclusive evidence. This indicates one can neither rule out sustained and relatively large effects nor close to zero effects of financial education at longer time spans due to a very limited number of studies that measure very long-run outcomes. We attribute the previous finding of a relatively rapid decay to the fact that Fernandes et al. (2014) chose to model this relationship in a meta-regression model with four covariate variables based on a sample of only 29 observations.[17] Thus, the evidence suggesting insignificant effects after time spans of more than 18 months is based on a very limited number of observations and should be viewed with caution in light of the large uncertainty around this estimated effect.

6 Discussion of the economic significance of financial education

[17] We also rerun their type of model (a regression of the estimated effect size on "linear effects of mean-centered number of hours of instructions, linear and quadratic effects of number of months between intervention and measurement of behavior, and the inter action of their linear effects" (Fernandes et al. 2014, p. 1867) with our updated data (419 observations within 52 studies) and find coefficient estimates with large standard errors (i.e., insignificant coefficients) throughout (see Table B6 in Appendix B).

As is true with any analysis of interventions, it is important to understand not just the statistical effect size but also the economic significance of the effects of financial education. A growing literature in education is concerned with interpreting effect sizes across studies, samples, interventions, and outcomes. This section discusses the choice in Fernandes et al. (2014) to focus on the "variance explained" as a measure of the effect size (Section 6.1). We couch our effect sizes into the recent literature on explaining and comparing the effects of education interventions (Section 6.2), provide a back of the envelope analysis of the cost-effectiveness of financial education interventions based on our findings (Section 6.3), and discuss the external validity of the RCT estimates by taking into account recent quasi-experimental studies (Section 6.4).

6.1 Statistical effect sizes

A main argument in Fernandes et al. (2014) is that even though the statistical effects of financial education on financial outcomes are positive in the overall sample, the magnitudes are small. However, Fernandes et al. (2014) create the illusion of miniscule effects (when, in fact, they can be economically significant) by using "variance explained," i.e., a squared correlation coefficient, as their effect size metric.

The fact that this metric creates the illusion of miniscule effects can be illustrated with a simple example. Consider the median effect of education (and specifically, structured pedagogy) interventions in developing countries, which is roughly 0.13 SD units (see Evans and Yuan 2019). Translating this to the (partial) correlation in Fernandes et al. (2014) results in a correlation coefficient of 0.06, which explains only 0.36 percent of the variance in learning outcomes. Thus, according to this criterion, this education intervention would be interpreted to be ineffective, as it "explains little of the variance." However, Evans and Yuan (2019) report that this is actually equivalent to a sizeable effect, approximately 0.6-0.9 years of "business as usual schooling," depending on their choice of specification. In further analysis, they estimate the returns to education (and specifically literacy) in Kenya, and estimate the net present value of this intervention to be 1,338 USD at an average annual income of 1,079 USD in 2015 PPP. Reported in this way, rather than the metric chosen by Fernandes et al (2014), these effects are unlikely to be considered economically miniscule. Thus, it can be problematic to rely upon the "variance explained" in determining the economic interpretation of statistical effect sizes.

6.2 Interpreting treatment effects in the education literature

Recent work in education interventions aims to compare effect sizes across heterogeneous treatments, populations, and outcomes—as we are doing in our analysis—and we turn to that work to get some guidance on interpreting effects. Kraft (2019) suggests five key considerations in determining whether or not programs are effective. First, one should make sure only studies with a causal interpretation (e.g., RCTs) are included in "effect sizes." Second, one should expect effects to be larger when the outcome is easier to change; this is particularly relevant if the intervention is *designed* to change the specific outcome. Third, one should take into account heterogeneous effects on different populations. Fourth, one should always consider costs per participant. A small effect size can have a large return on investment if the per participant cost is low. Fifth, one should consider whether the program is easily scalable. We have followed these recommendations.

With these five points in mind, Kraft (2019) further points to a scheme for assessing the effect of education interventions with academic outcomes (i.e., test scores) as the main outcome of interest. He suggests that effects larger than 0.20 standard deviations are "large," effects between 0.05 and 0.20 standard deviations are "medium," and effects under 0.05 standard deviations are "small." This classification is roughly consistent with the What Works Clearinghouse (2014), Hedges and Hedberg (2007) and Bloom et al. (2008). Our effects on financial knowledge in Figure 4 show an effect size of roughly 0.203, consistent with a "large" effect of an education intervention on test scores.

Kraft (2019) also notes that it is more difficult to affect long-run outcomes that are not directly addressed in the intervention. It is, thus, not surprising that effects on financial behavior are more modest than effects on financial knowledge. Even so, these effects are classified as "medium" in magnitude in his interpretation of effect sizes realized in RCTs.

6.3 Cost-effectiveness

While understanding effect sizes in standard deviation units is more consistent across educational interventions and more intuitive than "variance explained," a discussion of effect sizes is incomplete without quantifying costs, as also noted in Kraft (2019). Unfortunately, only 20 papers within the 76 studied include a discussion of cost. If we conduct a meta-analysis with only these papers, we find that the estimated treatment effects are smaller in the set of studies reporting costs than in the fully aggregated sample. In Appendix B Figure B6, we regress a binary indictor of reporting costs on sample and experiment characteristics to examine which are the studies that do report costs. The only notable difference is that studies reporting costs are more likely to involve low-income samples. Since we see no difference in effect sizes based on whether or not the intervention was targeted to low-income populations, we cannot precisely say what is driving the difference in effect sizes with respect to studies reporting costs.

To give readers a visual assessment of costs and effect sizes, we report the average costs by study in Appendix A Table A1 in 2019 U.S. dollars. Averaging across all studies reporting costs, the mean and median per participant costs are $60.40 and $22.90, respectively. Using the Kraft (2019) scheme with respect to effect sizes, an average cost of $60 per participant would be classified as a "low cost" educational intervention. It could be that studies reporting costs have, on average, lower costs than those that do not report costs. If that is the case, costs are understated, as are benefits since effect sizes are smaller in the reporting sample. Several studies mention their interventions had “minimal costs” but do not report a number; we do not include these studies in the cost estimates. Some programs may have costs that are difficult to quantify. Other programs may be difficult to scale. For example, Calderone et al. (2018) report a $25 per person cost and $39 per person benefit for a financial education program in India. However, they state the program is still too costly for a large company to implement at scale. While some studies pass a cost-benefit analysis on the surface, there may be other barriers prohibiting implementation.

Overall, our cost-effectiveness ratio is $60.40 per person for one-fifth of a standard deviation improvement in outcomes. Figure 5 displays the cost and effect size by outcome domain for each study. There are two direct takeaways from the figure. First, most effect sizes lie above the zero line but below 0.5 standard deviations. The effects below the zero line largely reflect papers that study the impact of financial education on remittances (e.g., switching to a cheaper financial product when transferring money across countries). Second, there does not appear to be a linear relationship between costs and effect sizes. Figure B7 in Appendix B displays the effect sizes and costs for each outcome domain separately, where we also include 95% confidence intervals for each estimate.

< Figure 5 about here >

To make the discussion more salient, we use one paper that clearly spells out the costs, from a large-scale randomized control trial in Peruvian schools (Frisancho 2018). That paper reports a cost per pupil of $4.80 USD and that a $1 increase in spending on the program yields a 3.3 point improvement in the PISA financial literacy assessment. Since this study represents financial education within a year-long class and average and median interventions in the sample are only 12 and 7 hours, respectively, it is likely that the average effect across studies corresponds to lower costs. Frisancho (2018) also shows that the course does not detract from performance in other courses, limiting opportunity costs.

Our back of the envelope estimate is conservative in that it does not consider positive externalities of the program. For example, Frisancho (2018) documents that in addition to improving student outcomes, teachers' financial literacy and credit scores also increase. Further, Bruhn et al. (2016) document positive "trickle up" effects for parents. Thus, financial education programs may have externalities beyond the target group, such as affecting behaviors of teachers, parents, and possibly peers (Haliassos et al. 2019).

6.4 External validity

While a benefit of only including RCTs is that there is little debate regarding their internal validity, it is more common to study long-term effects in quasi-experimental settings. There exists mounting quasi-experimental evidence that requiring U.S. high school students to complete financial education prior to graduating improves long-term financial behaviors. This body of literature uses a difference-in-difference strategy comparing students who would have graduated just before and just after the requirement was in place within a state with a requirement, as well as across states with and without requirements over the same time period.[18]

High school personal finance graduation requirements, which include standalone courses and personal finance standards incorporated into another required class or curriculum, show that financial education reduces non-student debt (Brown et al. 2016), increases credit scores (Brown et al. 2016; Urban et al. 2018), reduces default rates (Brown et al. 2016; Urban et al. 2018), shifts student loan borrowing from high-interest to low-interest methods (Stoddard

[18] Cole, Paulson, and Shastry (2016) used this method but studied "personal finance mandates" between 1957-1982, which often did not comprise course requirements but instead brought a representative from a bank to give a one-off lecture. The authors documented no effects of the education on investment or credit management behaviors. This was in contrast to Bernheim, Garret, and Maki (2001), who found that these same mandates improved investment behaviors, though they did not include state-level fixed effects in their analysis.

and Urban 2019), increases student loan repayment rates (Mangrum 2019), reduces payday loan borrowing for young adults (Harvey, 2019), and increases bank account ownership for those with only high school education (Harvey 2020). This recent literature as well confirms the findings in the meta-analysis.

7 Conclusions

Our analysis of the existing research on financial education using the most rigorous evaluation methods has three main findings.

First, financial education treatment effects from RCTs have, on average, positive effects on financial knowledge and behaviors. This result is very robust: it holds up to accounting for publication bias, including only adequately powered studies, looking only at studies published in top economics journals, and accounting for heterogeneity across studies. Financial education interventions have sizable effects on both financial knowledge (+0.2 SD units) and financial behaviors (+0.1 SD units). Thus, the treatment effects on financial knowledge are quite similar to or even larger in magnitude than the average effect sizes realized by educational interventions in other domains such as math and reading (see Hill et al. 2008; Cheung and Slavin 2016; Fryer 2016; Kraft 2018) and the effect sizes on financial behaviors are comparable to those realized in behavior-change interventions in the health domain (e.g., Rooney and Murray 1996; Portnoy et al. 2008; Noar et al. 2007) or behavior-change interventions aimed at fostering energy conserving behavior (e.g., Karlin et al. 2015). Our findings are in stark contrast to the findings presented in the first meta-analysis of the financial education literature (Fernandes et al. 2014). How can we interpret these differences in findings? While we are unable to replicate the original result on RCTs presented in Fernandes et al. (2014) (see Appendix D), we observe that the number of recent RCTs added to the database is driving the more positive result of financial education treatment effects on financial knowledge and behaviors. Additionally, we show that explicitly accounting for heterogeneity in studies and programs is crucial in assessing the average impact of financial education.

Second, there is no evidence to support or refute decay of financial education treatment effects six months or more after the intervention. Since only six studies in our sample look at impacts 24 months beyond the intervention, we cannot rule out that this effect is statistically different from short-run effects. Because the present literature is characterized by very few longer-term impact assessments, the evidence on the sustainability of effects is inconclusive. What we can say, however, is that we do not find evidence for dramatic decay up to six months after the intervention.

Third, we document that the estimates of statistical effect sizes are economically significant. We further document that many of the financial education interventions studied in randomized experiments are cost-effective. This finding is crucial, since the discussion of the effectiveness of financial education has focused on statistical effect sizes without considering their economic interpretation.

The evidence in this meta-analysis summarizes financial education interventions from 33 countries and six continents, across the lifespan of individuals. The analysis carefully accounts for heterogeneity across interventions. However, there are still some limitations. Since few RCTs study long-run effects, it is hard to determine the long-run impacts of these interventions. The same is true for the quality of the data used to study changes in financial behaviors: Few studies are able to link their experiments to administrative data, so the usual caveats of having to rely on self-reported survey data also apply to this literature. Future research should aim to collect longer-run administrative data or follow up with original participants from earlier field experiments. Finally, we encourage more studies to report on the costs of their programs, in order to provide policymakers with an estimate of cost-effectiveness.

References

Beuermann, D. W. and C. K. Jackson (2018). The short and long-run effects of attending the schools that parents prefer. Working paper. https://works.bepress.com/c_kirabo_jackson/37/

Bloom, H. S., Hill, C. J., Black, A. R., and Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1(4): 289–328.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley. http://dx.doi.org/ 10.1002/9780470743386

Bruhn, M., de Souza Leao, L., 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): 256–295.

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.

Cheung, A. and Slavin, R. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5): 283–292

Collins, J. M. and O'Rourke, C. M. (2010). Financial education and counseling - still holding promise. *Journal of Consumer Affairs*, 44 (3): 483–98.

Evans, D. and Yuan, F. (2019). Equivalent years of schooling: A metric to communicate learning gains in concrete terms. *World Bank Policy Research Working Paper No. 8752.*

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.

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

Fryer, R. G. (2016). The production of human capital in developed countries: evidence from 196 randomized field experiments. *NBER Working Paper No. 22130.*

Haliassos, M., Jansson, T., and Karabulut, Y. (2019). Financial literacy externalities. *Review of Financial Studies* 33 (2): 950–989.

Harbord, R. M., Higgins, J. P., et al. (2008). Meta-regression in Stata. *Stata Journal*, 8(4):493519.

Harvey, M. (2019). Impact of financial education mandates on young consumers' use of alternative financial services. *Journal of Consumer Affairs,* forthcoming.

Harvey, M. (2020). Does state-mandated high school financial education affect savings by low-income households? Working paper. https://static1.squarespace.com/static/5c4d314bb27e3999d515a9e4/t/5e0a1b2841180e2960023175/1577720633380/Harvey_FinEd_Savings_Working+Paper_v20191230.pdf

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.

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

Hedges, L. V. and E. C. Hedberg (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1): 60–87.

Hill, C. J., Bloom, H. S., Black, A. R., and Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3): 172–177.

Kaiser, T. and Menkhoff, L. (2017). Does financial education impact financial behavior, and if so, when? *World Bank Economic Review*, 31(3): 611–630.

Kaiser, T. and Menkhoff, L. (2019). Financial education in schools: A meta-analysis of experimental studies. *Economics of Education Review*. https://doi.org/10.1016/j.econedurev.2019.101930

Karlin, B., Zinger, J. F., and Ford, R. (2015). The effects of feedback on energy conservation: a meta-analysis. *Psychological Bulletin* 141(6): 1205–1227.

Kraft, M. A. (2019). Interpreting effect sizes of education interventions. *Educational Researcher*, forthcoming.

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

Lührmann, M., Serra-Garcia, M., and Winter, J. (2018). The impact of financial education on adolescents' intertemporal choices. *American Economic Journal: Economic Policy*, 10(3): 309–332.

Lusardi, A. and Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1): 5–44.

Lusardi, A., Michaud, P.-C., and Mitchell, O. S. (2017). Optimal financial knowledge and wealth inequality. *Journal of Political Economy*, 125(2): 431–477.

Mangrum, D. (2019). Personal finance education mandates and student loan repayment. Working paper. https://www.danielmangrum.com/research.html

Meager, R. (2019). Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics*, 11(1): 57–91.

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.

Noar, S. M., Benac, C. N., and Harris, M. S. (2007). Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. *Psychological Bulletin*, 133(4): 673–693.

OECD (2015). National strategies for financial education. OECD/INFE policy handbook, https://www.oecd.org/finance/National-Strategies-Financial-Education-Policy-Handbook.pdf.

Portnoy, D. B., Scott-Sheldon, L. A., Johnson, B. T., and Carey, M. P. (2008). Computer-delivered interventions for health promotion and behavioral risk reduction: A meta-analysis of 75 randomized controlled trials. *Preventive Medicine,* 47(1): 3–16.

Rooney, B. L. and Murray, D. M. (1996). A meta-analysis of smoking prevention programs after adjustment for errors in the unit of analysis. *Health Education Quarterly,* 23(1): 48–64.

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

Stoddard, C. and Urban, C. (2019) The effects of financial education graduation requirements on postsecondary financing decisions. *Journal of Money, Credit, and Banking,* forthcoming.

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.

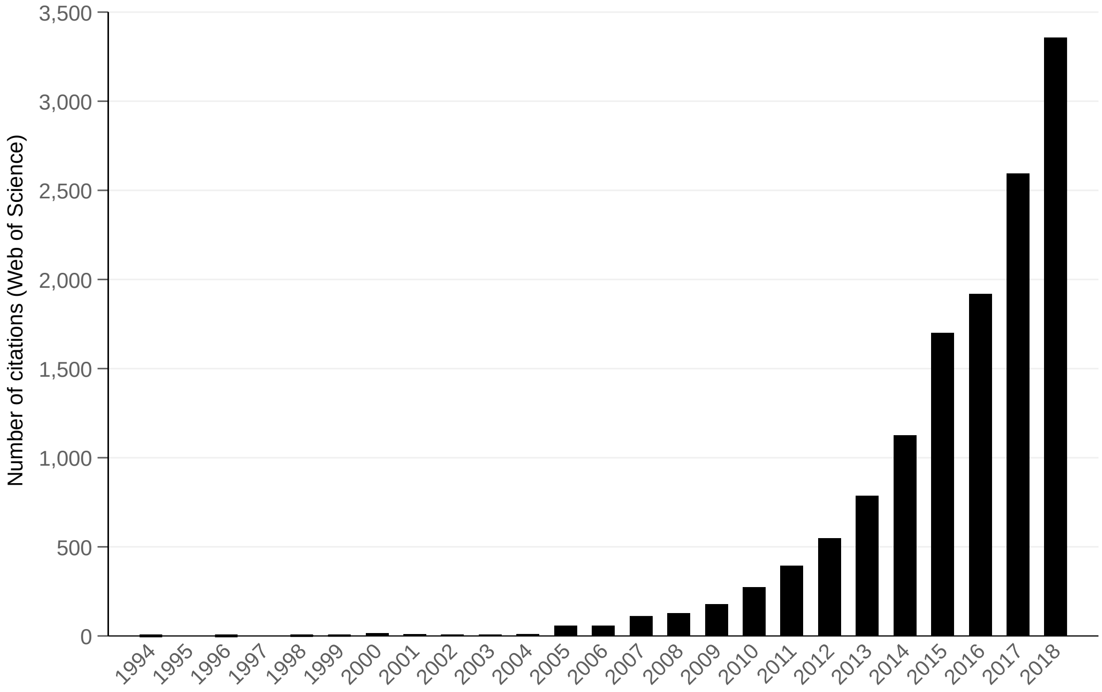
Tanner-Smith, E. E., Tipton, E., and Polanin, J. R. (2016). Handling complex meta- analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, 2(1): 85–112.

Urban, C., Schmeiser, M., Collins, J. M., and Brown, A. (2018). The effects of high school personal financial education policies on financial behavior. *Economics of Education Review*, forthcoming.

What Works Clearinghouse. (2014). WWC procedures and standards handbook (Version 3.0). *U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse.*

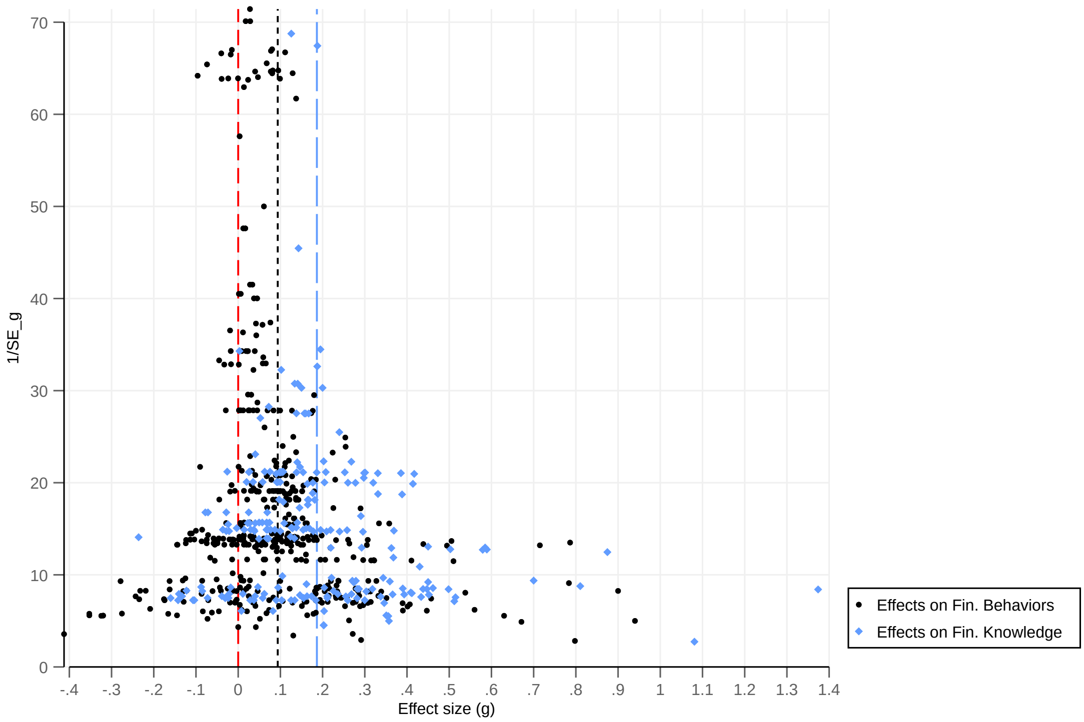
Xu, L., and Zia, B. (2012). Financial literacy around the world: An overview of the evidence with practical suggestions for the way forward. *World Bank Policy Research Working Paper No. 6107.*

Figure 1: Citations in the SSCI to the term "financial literacy" per year



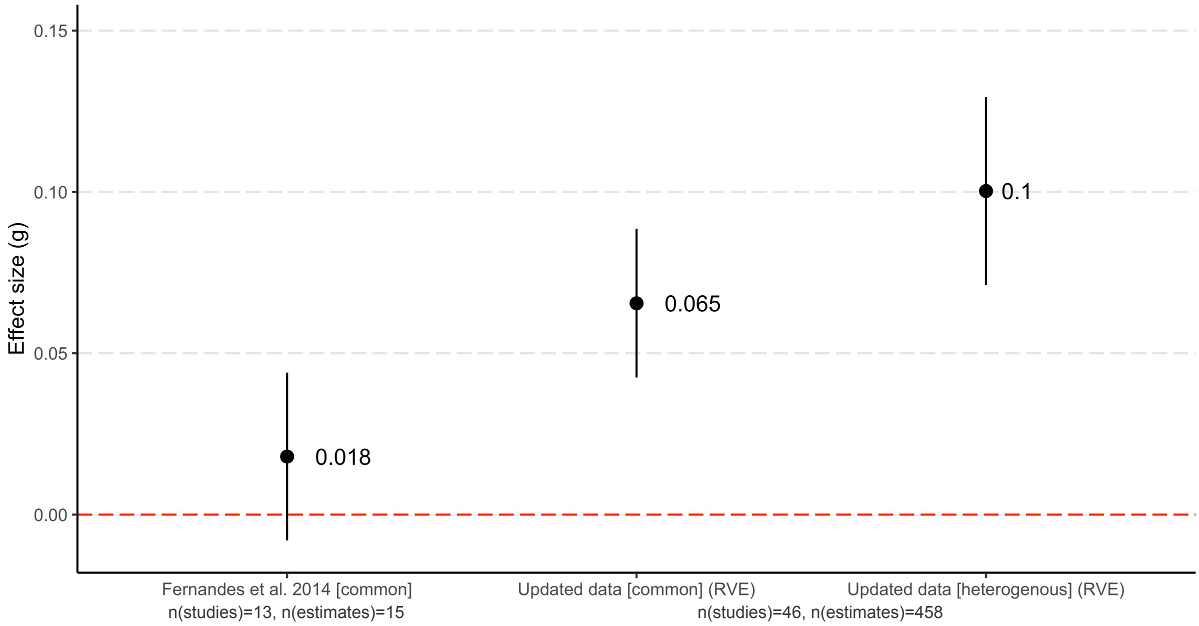
Notes: Number of citations within the social science citation index (Web of Science) to articles including the term "financial literacy" in the title or the abstract. Data from October 11, 2019.

Figure 2: Distribution of raw financial education treatment effects and their standard errors



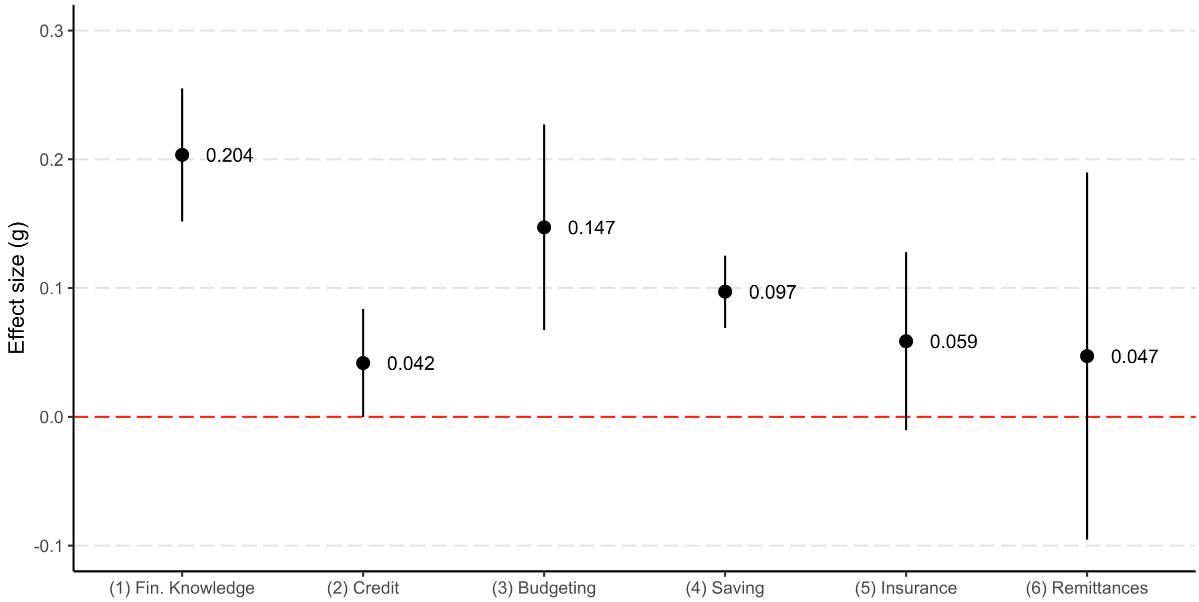
Notes: Effect size (g) is the bias corrected standardized mean difference (Hedges' g). 1/SE_g is its inverse standard error (precision). The number of observations in the treatment effects on financial behaviors sample is 458 effect size estimates from 64 studies. The number of observations in the treatment effects on financial knowledge sample is 215 effect size estimates from 50 studies. Thirty-eight studies report treatment effects on both types of outcomes. The mean effect size on financial behaviors is 0.0937 SD units, and the mean effect size on financial knowledge is 0.186 SD units.

Figure 3: Estimating the average effect of financial education treatment on financial behaviors in RCTs



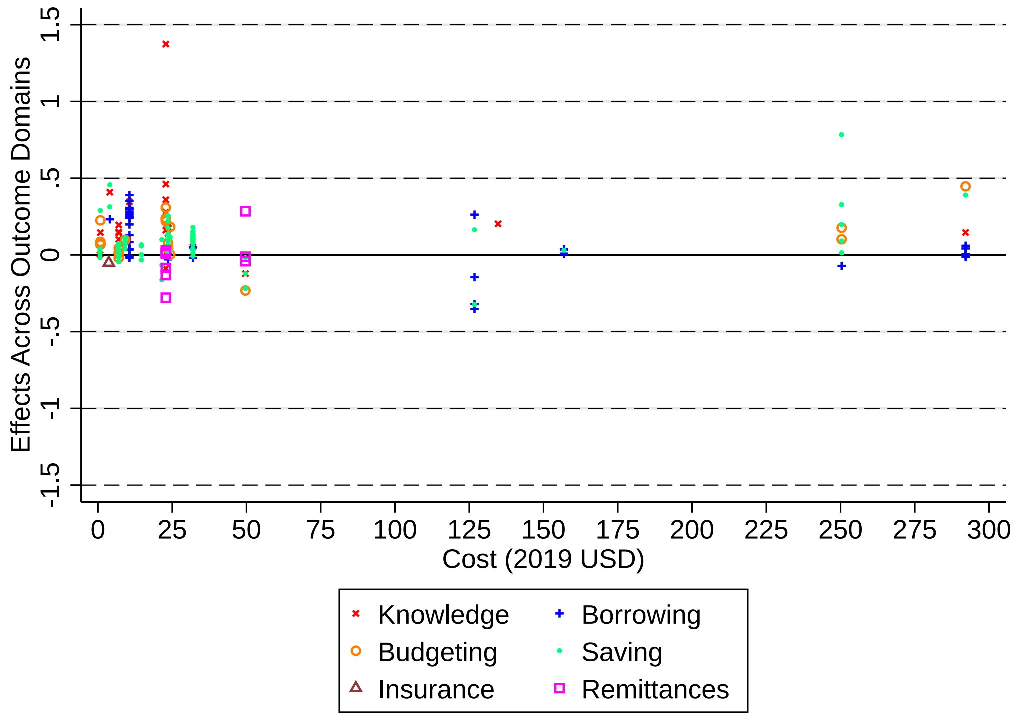
Notes: Fernandes et al. (2014) report weighted least squares estimates with inverse variance weights (common effect assumption). The results with updated data are from robust variance estimation in meta-regression with dependent effect size estimates (RVE) (Hedges et al. 2010) with $\tau^2 = 0$ in the common effect case, and τ^2estimated via methods of moments in the heterogeneous effects case. Fernandes et al. (2014) use within-study average effects and estimate the weighted average effect across 15 observations using inverse variance weights. Our estimates with updated data are based on multiple effect sizes per study and account for the statistical dependency (estimates within studies) by relying on robust variance estimation in meta-regression with dependent effect size estimates (Hedges et al. 2010). Dots show the point estimate, and the solid lines indicate the 95% confidence interval.

Figure 4: Financial education treatment effects by outcome domain



Notes: Results from robust variance estimation in meta-regression with dependent effect size estimates (RVE) (Hedges et al. 2010). The number of observations for the financial knowledge sample (1) is 215 effect size estimates within 50 studies. The number of observations for the credit behavior sample (2) is 115 within 22 studies. The number of effect size estimates for the budgeting behavior sample (3) is 55 within 23 studies. The number of observations in the saving behavior (4) sample is 253 effect size estimates within 54 studies. The number of observations in the insurance behavior sample (5) is 18 effect sizes within six studies. The number of observations on remittance behavior (6) is 17 effect size estimates reported within six studies. Dots show the point estimate, and the solid lines indicate the 95% confidence interval.

Figure 5: Cost of intervention and effect sizes



Notes: The graph depicts the cost and effect sizes for each outcome domain among the 20 experiments that report costs. Each data point is an effect size for an outcome studied. Figure B7 in Appendix B provides a graph for each outcome domain that contains standard errors of the estimates.

Table 1: Descriptive statistics

Variable	Obs.	Mean	Median	Std. Dev.	Min.	Max.
Hedges' *g*	677	0.123	0.098	0.183	-0.413	1.374
SE (*g*)	677	0.084	0.072	0.049	0.007	0.365
Time span (in weeks)	643	30.239	25.800	31.537	0.000	143.550
Intensity (in hours)	604	11.709	7.000	16.267	0.008	108.000
Mean age (in years)	650	33.480	38.300	12.480	8.500	55.000
Children (< age 14)	677	0.075	-	-	0.000	1.000
Youth (age 14 to 25)	677	0.201	-	-	0.000	1.000
Adults (> age 25)	677	0.724	-	-	0.000	1.000
Low income (yes=1)	677	0.725	-	-	0.000	1.000
Developing economy (yes=1)	677	0.604	-	-	0.000	1.000
Top econ journal (yes=1)	677	0.267	-	-	0.000	1.000

Note: Descriptive statistics at the estimate-level, i.e. we consider the total of 677 effects reported in 76 RCTs.

Table 2: Financial education treatment effects by subgroups of studies and populations

Subgroup	Effect size (g)	SE	95% CI Lower bound	95% CI Upper bound	n(Studies)	n(effects)
Panel A: Treatment effects on *financial behaviors*						
(a) By country income						
High income economies	0.1127	0.0316	0.0478	0.1777	32	129
Developing economies	0.0928	0.0130	0.0660	0.1195	32	329
(b) By respondent income						
Low income individuals	0.0993	0.0194	0.0600	0.1387	43	367
General population	0.1035	0.0219	0.0571	0.1500	21	91
(c) By age of participants						
Children (< age 14)	0.0640	0.0186	0.0188	0.1091	9	36
Youth (age 14 to 25)	0.1203	0.0415	0.0250	0.2155	11	92
Adults (> age 25)	0.1068	0.0205	0.0653	0.1483	44	330
(d) By type of publication						
Top econ. journals	0.0833	0.0235	0.0325	0.1342	15	161
Other publications	0.1075	0.0183	0.0704	0.1445	49	297
(e) By delay between treatment and measurement of outcomes						
Delay of < 6 months	0.0991	0.0169	0.0645	0.1337	34	180
Delay of ≥ 6 months	0.0710	0.0137	0.0425	0.0995	28	260
Delay of ≥ 12 months	0.0878	0.0200	0.0450	0.1308	18	134
Delay of ≥ 18 months	0.0653	0.0192	0.0209	0.1098	10	49
Delay of ≥ 24 months	0.0574	0.0225	0.0013	0.1136	7	32
Panel B: Treatment effects on *financial knowledge*						
(a) By country income						
High income economies	0.2591	0.0415	0.1738	0.3443	29	135
Developing economies	0.1392	0.0218	0.0934	0.1851	21	80
(b) By respondent income						
Low income individuals	0.2238	0.0395	0.1428	0.3049	30	120
General population	0.1835	0.0310	0.1183	0.2486	20	95
(c) By age of participants						
Children (< age 14)	0.2763	0.1098	0.0076	0.5450	7	15
Youth (age 14 to 25)	0.1859	0.0390	0.1015	0.2703	16	40
Adults (> age 25)	0.2001	0.0282	0.1418	0.2583	28	160
(d) By type of publication						
Top econ. journals	0.1572	0.0379	0.0648	0.2497	8	46
Other publications	0.2142	0.0299	0.1537	0.2746	42	169
(e) By delay between treatment and measurement of outcomes						
Delay of < 6 months	0.2305	0.0319	0.1654	0.2956	36	142
Delay of ≥ 6 months	0.1408	0.0289	0.0775	0.2041	15	57
Delay of ≥ 12 months	0.1406	0.0367	0.0166	0.2646	5	5
Delay of ≥ 18 months	-	-	-	-	0	0
Delay of ≥ 24 months	-	-	-	-	0	0

Notes: This table reports average effects of financial education treatment on financial behaviors (Panel A) and financial knowledge (Panel B) estimated via RVE. Ten studies with 34 effect size estimates are missing information about the delay between treatment and measurement of outcomes.