Journal of Economic Behavior and Organization 214 (2023) 561–573

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Are retirement planning tools substitutes or complements to financial capability?

Gopi Shah Goda [a], Matthew R. Levy [b], Colleen Flaherty Manchester [c], Aaron Sojourner [d], Joshua Tasoff [e,*], Jiusi Xiao [e]

[a] *Stanford University, 366 Galvez Street, Stanford, CA 94305, United States of America*
[b] *London School of Economics, Houghton Street, London WC2A 2AE, UK*
[c] *University of Minnesota, 321 19th Avenue South, Minneapolis, MN 55455, United States of America*
[d] *W.E. Upjohn Institute for Employment Research, 300 S. Westnedge Ave, Kalamazoo, MI 49007, United States of America*
[e] *Claremont Graduate University, 160 East Tenth St., Claremont, CA 91711, United States of America*

ARTICLE INFO

JEL classification:
D14
G53
J32

Keywords:
Retirement planning
Retirement saving
Exponential-growth bias
Present bias
Financial literacy
Financial capability

ABSTRACT

We conduct a randomized controlled trial to understand how a web-based retirement saving calculator affects workers' retirement-savings decisions. In both the treatment and active control conditions, the calculator projects workers' retirement income goal. In the treatment condition only, it also projects retirement income based on defined-contribution savings, prominently displays the gap between projected goal and actual retirement income, and allows users to interactively explore how alternative, future contribution choices would affect the gap. The treatment increased average annual retirement contributions by $174 (2.3 percent). However, effects were larger for those with higher measures of financial knowledge, suggesting this type of tool complements, rather than substitutes for, underlying financial capability.

Acknowledgments The preanalysis plan for this project was registered at the Social Science Registry AEARCTR-0002129. Human subjects protocol was determined as exempt by Stanford IRB (#38948), University of Minnesota IRB (#1607E90001) and Claremont Graduate University IRB (#2813). This research was supported by the U.S. Social Security Administration through grant #5-RRC08098400-10 to the National Bureau of Economic Research as part of the SSA Retirement Research Consortium and the Laura and John Arnold Foundation through a grant to the Stanford Institute for Economic Policy Research (SIEPR). The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the federal government, the NBER, the Laura and John Arnold Foundation, SIEPR, or any other institution with which the authors are affiliated. We are grateful to Paula Gablenz and Konhee Chang for exceptional research assistance.

* Corresponding author.
E-mail addresses: gopi@stanford.edu (G.S. Goda), m.r.levy@lse.ac.uk (M.R. Levy), cmanch@umn.edu (C. Flaherty Manchester), sojourner@upjohn.org (A. Sojourner), joshua.tasoff@cgu.edu (J. Tasoff), jiusi.xiao@cgu.edu (J. Xiao).

https://doi.org/10.1016/j.jebo.2023.08.001
Received 18 August 2022; Received in revised form 2 August 2023; Accepted 4 August 2023

Available online 30 August 2023

1. Introduction

Determining how much to save for retirement is a complex problem that, in the era of defined contribution (DC) retirement saving plans, largely falls on the individual. Solving for one's optimal retirement saving contribution in a given year requires simultaneously setting a target income in retirement and determining what contribution path enables one to meet that goal, taking into account investment returns, expected retirement age, and other sources of retirement income, such as Social Security and defined benefit pension income. Navigating this problem is challenging due to its high dimensionality, considerable uncertainty, and the limited opportunity to learn from mistakes.

There is reason to believe that many people are not well-equipped to solve this complex problem. Rates of understanding for basic financial concepts are low (Lusardi and Mitchell, 2014). Further, limited financial understanding is one explanation for the disproportionate influence of defaults—which dictate employee outcomes when no choice is made—on participation and contribution decisions (Madrian and Shea, 2001; Beshears et al., 2009). Individuals may look to default settings as implicit saving advice, yet such settings may not be aligned with individuals' retirement lifestyle goals. In addition, Goda et al. (2019) show that low financial literacy and lack of understanding of exponential-growth bias are associated with lower retirement wealth accumulation among retirement-age individuals. These findings underlie public policy concern regarding the extent to which low financial capability fuels the limited retirement savings observed among many individuals.

Plan sponsors and academic researchers have sought to improve retirement saving decisions by supporting employee decision-making through information campaigns and via online retirement saving tools or calculators. Three key factors that determine whether these informational interventions are likely to be successful in addressing inadequate retirement saving are 1) who selects into using them, 2) how they affect contribution behavior among those who use the tool, and 3) how the intervention differentially affects financially more vulnerable populations. Are such tools effective at raising financial decision-making capacity across the board, including those with limited financial literacy, or do the tools themselves require a sufficient understanding of financial concepts in order to be effective? That is, are such tools a substitute for existing financial knowledge or a complement? Often plan sponsors introduce decision-support tools with the goal of increasing participation among those with lower financial knowledge, and thus implicitly assume these tools are substitutes.

To address these questions, we conduct a randomized controlled trial among employees at the U.S. Office of Personnel Management (OPM), an agency of the federal government. Federal employees have an employer-sponsored retirement savings program similar to a 401(k), called the Thrift Savings Plan (TSP), in which agencies match employee contributions. Our design randomly assigns employees to receive one of two online retirement saving tools: a treatment and an active control. Both tools elicit information on the participant's desired lifestyle in retirement, current earnings, and expected retirement age in order to display their target retirement income as well as collect information on inputs to a retirement income projection. The tools differ in how complete this projected income calculation is. The "treatment tool" incorporates expected social security, federal defined-benefit pension income, and existing TSP savings and contribution levels into the projection, allowing participants to see whether the projection aligns with their target and dynamically assess how their TSP contributions map to their retirement income. In contrast, the "active control tool" omits retirement income stemming from TSP in the income projection. Instead, participants in the active control are asked to make their own assessment as to how much additional retirement income their accumulated TSP savings and future contributions will provide to assess whether they are on track to meet their retirement income target.

The difference between these two conditions isolates the effect of financial decision-making involving computations that map current contribution behavior to financial resources in retirement. The accuracy of the subjective mapping used by a person may vary with their financial capability and willingness to engage in effortful thinking and planning. Past research has specifically implicated exponential-growth bias, present bias, and financial illiteracy as attributes implicated in low retirement savings (Goda et al., 2014; Brown and Previtero, 2014; Goda et al., 2019; Lusardi and Mitchell, 2011a). The additional information provided in the treatment condition removes the need to make exponential computations, which require effort that is prone to postponement by present-biased individuals. Therefore, our treatment is designed specifically to overcome the exponential-growth bias and present bias that could otherwise lead to suboptimal decision-making.

We find that approximately half of employees (48 percent) select into using the tool, and selection is correlated with pre-intervention TSP contributions. We then evaluate whether the treatment tool affected TSP contributions relative to the active control, and how the effect varied across employees using a treatment-on-the-treated (TOT) estimation approach. Overall, we find that the treatment increased average annual contributions by \$174 among those who used it relative to those using the control tool. We examine heterogeneous treatment effects across multiple measures of financial capabilities. We find that the treatment effect is significantly greater for those with: higher measures of financial literacy, a college degree, and a higher financial-capability index score derived from factor analysis. We do not find evidence that exponential-growth bias, present bias, preintervention contributions, or other factors derived from factor analysis significantly predict the treatment effect.

Our study relates to several strands of literature. First, there is extensive evidence documenting the effects of a wide variety of retirement saving interventions, such as automatic enrollment (Madrian and Shea, 2001; Choi et al., 2004), retirement income projections (Goda et al., 2014; Dolls et al., 2018), commitment devices (Thaler and Benartzi, 2004), peer information (Duflo and Saez, 2003; Beshears et al., 2015), reducing complexity (Choi et al., 2006; Beshears et al., 2013) and anchoring (Choi et al., 2017). In general, this literature shows that small perturbations in choice architecture can have large effects on retirement saving decisions, providing indirect evidence that there are barriers present that keep people from optimizing retirement saving decisions.

Second, there is a large literature that evaluates interventions designed to address low financial literacy (e.g., Bernheim et al., 2001; Bernheim and Garrett, 2003; Lusardi, 2008; Mandell and Klein, 2009; Gale and Levine, 2011; Hastings et al., 2013; Austin and

Arnott-Hill, 2014). The general findings of this literature, summarized in a meta-analysis by Fernandes et al. (2014) and corroborated by Willis (2021), are that interventions to improve financial literacy explain a small amount of the variance in the financial behaviors studied, and that the effects are weaker in low-income samples. However, this literature suggests a role for "just-in-time" financial education that is tied to specific behaviors.

Finally, a broader literature investigates the selection into take-up of interventions in other contexts and includes several examples of these interventions reaching lower-need populations. For example, those who participate in workplace wellness programs tend to have lower medical expenditures and healthier behaviors than nonparticipants (Jones et al., 2019). Similarly, a randomized evaluation of an informational intervention to increase take-up of Supplemental Nutrition Assistance Program (SNAP) benefits found evidence that those who responded to the treatment had higher income and better health than the average enrollee in the control group (Finkelstein and Notowidigdo, 2019). There is also evidence that decision support tools may reach populations with less predicted response (Bundorf et al., 2019). Similarly, Hackethal et al. (2012) find that richer, older and more experienced investors are more likely to seek financial advisors in the German context, rather than being used by less informed or unsophisticated investors. These examples show how interventions designed to improve outcomes for those with greater need can instead widen disparities in outcomes.

Our study makes two main contributions to the literature. First, we find evidence that online retirement-income tools, specifically the part of the treatment focused on exponential computations, lead to modest increases in retirement contributions. These findings are similar to Goda et al. (2014) and Song (2020) who find evidence that providing information that corrects underlying bias in exponential growth can increase retirement savings in the United States and in China, but in contrast with findings in Olckers (2021) and Fuentes et al. (Forthcoming), which show negligible impacts of a retirement income calculator in South Africa and Chile. While course-based financial literacy interventions aimed at influencing financial decision-making over the life course have shown limited success, this "just-in-time" tool-based intervention delivers an immediate, albeit modest effect on a financial decision. Yet, due to the relatively high inertia in retirement elections, the effect on contributions may be long-lasting, leading to non-trivial increases in retirement wealth.

At the same time, we find that tool-based interventions have important limitations, including who selects into using the tool and whose decisions are more affected, which connects to this paper's other contribution. In particular, we find evidence that retirement-income projections are complements rather than substitutes to financial capability. The online tool delivery provides the ability to track tool users and link engagement with the tool to outcomes. Selection into treatment is higher among those with higher pre-intervention contributions, and the treatment effect is larger for those with higher measures of financial literacy and education. This finding is important as it suggests that retirement planning tools are unlikely to be sufficient to overcome biases and/or substitute for shortfalls in financial knowledge that may prevent optimal decision-making. In fact, such tools may be poised to widen wealth gaps between those with higher relative to lower capabilities for making financial decisions, similar to literature cited above in other contexts.

2. Experimental design and data

2.1. Retirement plan setting

As federal workers, OPM employees participate in a defined contribution plan known as Thrift Savings Plan (TSP), in which the employer makes a base contribution of 1 percent of pay and matches employee contributions up to 5 percent of pay.[1] Employees can contribute up to the IRS maximum each year, which was $18,000 in 2017. Employees are also covered by a defined benefit pension.[2] Employees may elect to invest their contributions in five different funds or a life-cycle option, which is a mix of the other funds based on the employee's age.

A 2015 TSP report indicates that approximately half of federal employees were not contributing enough to TSP to maximize the agency match (OPM, 2015). The proportion qualifying for the full match is even lower for recent hires, who are covered by a 3 percent automatic enrollment provision introduced in 2010. Concern about employees failing to maximize the match since automatic enrollment began prompted OPM leaders to seek an effective online retirement saving tool to improve TSP contribution decisions for federal employees.

2.2. Intervention

In partnership with OPM, we designed both a treatment and an active control version of a new online retirement saving tool with the aim of 1) providing employees with both a target retirement income and a projected retirement income, and 2) isolating the effect of translating their TSP asset level and any potential contribution stream into a projected retirement-income stream on outcomes. The tool rolled out in November 2017. The two versions of the tool—treatment and active control—were made as similar as possible except that the active control did not provide any information on how TSP balances and contributions translated into retirement income. This allows us to isolate the effect of the income projection from any other tool features. The tool begins by

[1] The agency matches dollar-for-dollar on an employee's contributions up to 3 percent of pay and $0.50 to the dollar for the next 2 percent of pay.

[2] Employees hired before 1984 are covered by a more comprehensive defined benefit plan and receive no base and no match on employee contributions to TSP, yet can contribute on their own up to the individual maximum.

asking the user a series of questions to determine their target income in retirement, such as their date of birth, when they started working for the federal government, their annual salary, their expected retirement age, and their desired lifestyle in retirement: 70, 85, 100, or 115 percent of current pay. Participants are visually shown the "goal" as a vertical bar, represented as a monthly annuitized income target for retirement. Then participants are asked questions to produce a projected retirement income based on their current assets and saving rate, including their TSP account balance, TSP annual contribution, pension coverage, and Social Security expectations. The main difference between the treatment and active control conditions is that the income projection of the former uses all information provided, while in the active control, it provides projections based only on pension and Social Security income, and states that retirement income from TSP is an additional amount on top of these other sources. Fig. 1 shows the difference between the treatment and active control conditions in terms of the core visual that compares the retirement income goal and the retirement income projection. Screenshots of the entire tool are available in Appendix G.

After displaying the projection, the treatment tool allows users to manipulate sliders to adjust TSP contributions to see how the projection changes relative to the goal in real time. Some parameter values of the economic environment are needed to create the income projection such as the income growth rate, inflation rate, and expected real rate of return. All calculation formulas are available in Appendix H. We provide default values for annual income growth and for inflation of 3 percent and 2 percent, respectively, which the user can modify. Because the real rate of return depends on one's retirement portfolio, there may be considerable variation in people's expected rate of return and desired lifestyles for retirement. We randomize these assumptions to test whether these default parameters affect saving behavior. The default rate of return is randomly assigned to 5 percent or 8 percent, and desired retirement lifestyle is randomized to 85 percent or 100 percent of income.[3] As with the other assumptions, all participants have the option to change these assumptions using sliders and can view how they change the income projection.

Both versions of the tool end by showing participants a printable summary of their current TSP contribution levels and a link to the TSP website and phone number with instructions on how to update contribution rates. The printable summary for the treatment tool also includes the last slider position for the TSP election.

Prior to the intervention, we surveyed the employees to measure background characteristics and behavioral parameters not present in administrative data. The survey was fielded between March 29, 2017, and April 14, 2017. OPM emailed each employee an initial survey invitation and two reminders to nonrespondents. Of the 5,426 employees, 1,435 completed the survey, a 26 percent completion rate. Through the survey, we measure financial capabilities, including exponential-growth bias (EGB), measures of financial literacy, and college degree completion. The survey also elicited time preferences, including the long-term discount rate and a measure of present-biased preferences. The full survey instrument is available in Appendix I.

We randomly assigned the 5,426 unique individuals employed at OPM in December 1, 2017, to have access with equal probability either to the treatment tool or the active control tool. We stratified participants based on survey completion with 50 percent of completers and 50 percent of noncompleters each getting the treatment condition. Within a survey-response group (completers/noncompleters) we stratified on total pay, age, TSP total amount, and gender. Survey completers were also stratified on their mean response to the EGB elicitation and mean response to the time-preference elicitation. OPM emailed each employee a personalized link to the appropriate version of the tool. Employees received an invitation to use the tool on December 1, 2017. Subsequent reminder emails were sent to those who had not yet clicked the link on December 7, December 18, and January 11. There was no differentiation in the invitation emails between the treatment and active control groups.

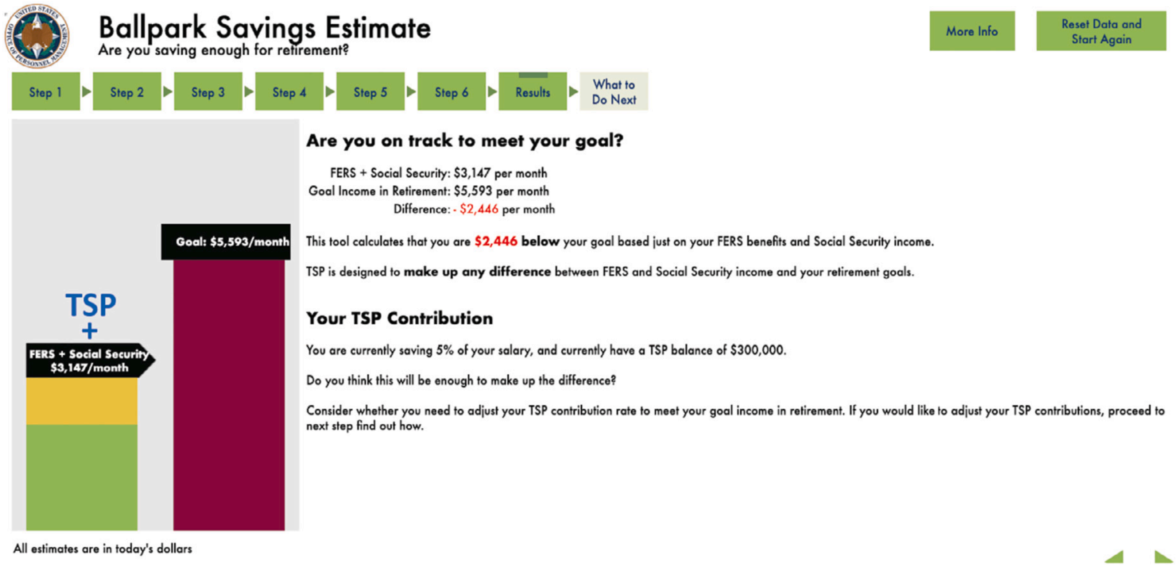
2.3. Data and analysis samples

Our data include each individual's monthly TSP contribution elections and demographic characteristics from administrative HR records from August 2014 to April 2018. We match these data with survey data collected in March and April of 2017, and data on whether each individual chose to use the assigned tool or not: 2,625 (48 percent) did.

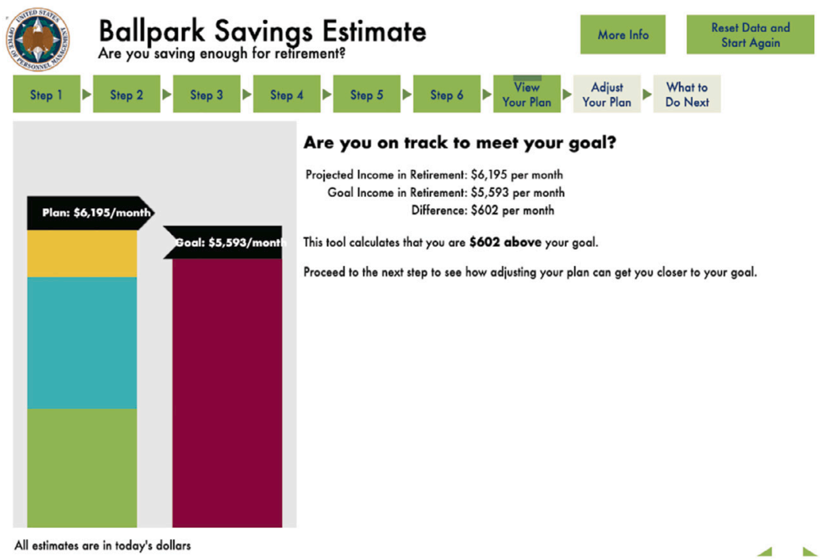
We use two analysis samples. The first relies just on administrative records, including TSP contributions and employee characteristics recorded in HR files. This sample consists of the 2,625 unique employees who used the tool and their 152,198 total individual-by-month contribution observations. The second analysis combines the TSP and HR records with survey responses and captures 1,435 unique individuals with 85,974 total individual-by-month observations. Appendix E presents a schematic of these samples.

We examine whether there were significant differences in baseline characteristics between individuals assigned to active control versus treatment (Appendix Table A.1). The joint test of null difference across baseline characteristics has a p-value of 0.96, reflecting successful random assignment. Appendix Table A.2 compares survey completers and noncompleters on administrative variables, which are fully observed. Survey completers are older, are more likely to be white, are higher-paid, and contribute more to TSP than noncompleters. To clarify which characteristics are most strongly associated with response conditional on the other characteristics, Appendix Table A.3 reports estimates from a logit model of survey response. In this model, many observable characteristics predict response, but importantly age, pay, and length of tenure do not.

[3] The results of these regressions are available in Appendix C in Tables C.11–C.13. The high values of these default parameters had an insignificant effect on savings relative to the low values.



(a) Active Control



(b) Treatment

Fig. 1. Screenshot of examples from each treatment condition.

2.4. *Survey measures*

We perform our primary heterogeneity analysis on the subsample who completed the survey. Below we describe our measures of financial capability, which are central for assessing heterogeneous treatment effects, and the elicitation of time preferences. Finally, we present findings from an exploratory factor analysis of the covariate space that shows the construction of a factor that aligns with measuring financial capability.

2.4.1. *Exponential-growth bias*

We hypothesize that exponential-growth bias plays an important role in creating a gap between individuals' ideal savings rate and their actual savings rate. Exponential-growth bias is the tendency to neglect compound interest (Stango and Zinman, 2009). Forecasting one's retirement savings without the use of a tool requires considerable sophistication. The lack of an accurate forecast along with exponential-growth bias may cause people to underestimate the benefits of saving for retirement. Because the intervention operates by explicitly computing the exponential growth of the user's savings (along with other computations), those with greater bias may benefit from the intervention more. More precisely, because undersaving is likely a larger problem than oversaving (see, for example, Goda et al., 2019, who show that exponential-growth bias is correlated with lower retirement savings), people with more exponential-growth bias may exhibit larger treatment effects if the tool compensates for their bias.

We use the parametric model of Levy and Tasoff (2016) given below.

$$p(\vec{r}, t; \alpha_i) = \prod_{s=t}^{T-1}(1 + \alpha_i r_s) + \sum_{s=t}^{T-1}(1 - \alpha_i) r_s \tag{1}$$

When $\alpha_i = 0$, the individual perceives growth to be linear, fully neglecting compound interest. When $\alpha_i = 1$, the person correctly perceives growth to be exponential. Values of $\alpha_i \in (0, 1)$ generate perceptions between linear and exponential growth. Values > 1 reflect overestimation of the returns to compounding. To measure exponential-growth bias, we include three hypothetical investment questions in our survey that ask for the value of an asset after a certain amount of time.[4] For each question k and each individual i, we construct a measure of exponential-growth misperception that minimizes the distance between the response and the correct answer informed by Equation (1) similarly to Goda et al. (2019). Performance on these questions by OPM employees was similar to the U.S. population: between 29 and 33 percent of survey participants answered the questions within 10 percent of the correct value as compared to 23 to 31 percent in a representative U.S. sample (Goda et al., 2019).

2.4.2. Time preferences

We hypothesize that present-biased individuals are more likely to have gaps between their ideal savings rate and actual savings rate due to procrastination. If so, displaying the gap may be a cue that inspires them to make a change. Though theory does not make a sharp prediction about the direction of change, we explore whether the treatment differentially affects participants based on the degree of their present bias.

We use a “time-staircase” procedure to construct a simple measure of present bias, which we refer to as “Beta,” as well as of the long-run discount factor (“Delta”) in an approach similar to Goda et al. (2019). The method was developed by Falk et al. (2016) for measuring only the long-run discount factor. Staircases have these forms:

Present-Future Staircase: Would you rather receive \$100 today or \$[X] in 12 months?
Future-Future Staircase: Would you rather receive \$120 in 12 months or \$[Y] in 24 months?

Subjects begin with a common value of [X] or [Y]. If a subject indicates they prefer the money sooner (later), then the second dollar amount increases (decreases) on the next question.[5] For each staircase, subjects answer five questions, gradually narrowing the interval that contains the indifference point. Since the questions are binary and have parallel structure, they are easily understood and can be answered quickly. Participants were asked these questions for a 12-month (as shown above) and a 6-month time interval, for a total of four sets. The complete staircase questions are presented in Appendix I. We randomize the order of the staircases and use different base values for the different sets of questions (i.e., the Present-Future Staircase always begins with \$100 today and the Future-Future Staircase with \$120 in 12 months) to minimize the influence of mechanical responses. While this staircase method did not involve real stakes, Falk et al. (2016) show that behavior between a no-stakes and real-stakes version is highly correlated.[6] From these staircases we construct measures of Beta and Delta from the implied indifference point.[7]

2.4.3. Measure of financial literacy

Employees with low financial literacy may struggle to make retirement savings decisions, due to not knowing what an appropriate savings rate is. In addition, low financial literacy may create difficulty regarding the process of implementing changes. We hypothesize that employees with low financial literacy would have bigger gaps between their ideal savings rate and their actual savings rate, and that the intervention will have larger treatment effects on those with low financial literacy if the savings tool serves as a substitute for financial capability.

We measure basic financial literacy using the five-item battery of financial literacy questions developed by Lusardi and Mitchell (2011b) and widely used since then (Lusardi and Mitchell, 2014). These questions measure understanding of inflation, diversification, compound interest, mortgage payments, and bond prices using multiple choice questions. We omit “don’t know” as a response option, in line with research that doing so significantly improves scores for female respondents (Bucher-Koenen, Tabea et al., 2017).[8] OPM employees performed well on these questions relative to the U.S. population; percent correct on each of the five questions ranged between 39 and 95 percent for OPM employees, compared to 21 and 70 percent for a representative sample of the U.S. population (Lusardi and Mitchell, 2011b). Similarly, the share of employees who answered all five questions correctly was 30 percent, relative

[4] An example question is, “An asset has an initial value of \$100 and grows at an interest rate of 10 percent each period. What is the value of the asset after 20 periods?” All three questions are available in Section 3 of Appendix I.

[5] In our survey instrument, the future value X was always greater than 100 and Y was always greater than 120.

[6] The authors find a correlation between the staircase measures and incentivized experimental measures of 0.524. This correlation is close to the test-retest correlation of 0.664 for the incentivized experiment.

[7] We cannot identify the indifference point for those who select the upper bound of the time staircase. In this case, we use the upper bound value plus the difference between that value and the second-to-last value to determine the indifference point. We include a dummy variable for those with these imputed values in the analysis.

[8] We thank an anonymous referee for pointing this out.

Table 1
Factor loading matrix.

Variable	Factor1 Demographics	Factor2 Seniority	Factor3 Financial Capability	Factor4 Time Preference	Factor5 Household Composition	Factor6 Hispanic Factor	Uniqueness
Age	-0.0753	0.6838	0.0146	0.0648	-0.2091	-0.07	0.4738
Male	0.2269	-0.0046	0.3806	0.046	0.5064	0.0223	0.5446
Years of Schooling	-0.0993	-0.1911	0.7269	-0.0084	-0.1586	0.1145	0.3869
Race = White	0.925	-0.0198	-0.0022	0.0105	-0.0082	-0.2718	0.0699
Race = Hispanic	-0.0756	-0.0451	0.024	0.0178	-0.025	0.9097	0.1632
Race = Black	-0.9478	0.0585	-0.0297	-0.0367	-0.0067	-0.1584	0.071
Household Size	-0.0492	-0.0578	-0.0828	-0.0419	0.8686	-0.0349	0.2299
Tenure(in years)	-0.0802	0.8116	-0.131	0.0262	0.063	-0.0457	0.311
Is Supervisor	0.0577	0.4178	0.3047	-0.0493	0.2453	0.2889	0.5832
Tenure Description = Permanent	-0.0107	0.6444	-0.02	-0.0151	-0.0988	-0.012	0.5741
Std. Alpha	0.0448	0.1002	0.349	-0.0211	0.0972	-0.3106	0.7598
Std. Beta	0.0349	-0.0148	-0.0841	0.8349	-0.074	-0.0388	0.2875
Beta-Delta	0.0313	0.0673	0.1772	0.7921	0.0388	0.0725	0.3289
Financial Literacy	0.1299	0.0207	0.7042	0.1154	0.0648	-0.0656	0.4649
Eigenvalue	2.07686	1.75206	1.50360	1.31937	1.05755	1.04191	

Notes: The principal component analysis generated 14 factors, but factors with eigenvalue greater than 1 are retained and reported. Table 1 reports the rotated factor loading matrix from the principal component analysis for the retained factors. Parallel analysis is performed, as shown in Appendix F.

to 10 percent for the U.S. population, suggesting that OPM employees are more financially literate than average. In our subsequent analysis, we use a z-score of the financial literacy measure standardized within the sample.

2.4.4. Factor analysis

To understand heterogeneity in treatment effects, we take two approaches. First, we look for heterogeneity along theoretically important dimensions—such as financial literacy, exponential-growth bias (EGB), present bias (beta), educational attainment, prior TSP contribution levels—one at a time. Second, we pool information across multiple measures of financial capability and reduce dimensionality by estimating a latent factor and looking at heterogeneous treatment effects between individuals with more or less of this factor.

For the second approach, the first step is to reduce the dimensionality of the heterogeneity by conducting a principal component analysis of the baseline characteristics. Specifically, we include age in years, gender, years of schooling, race/ethnicity categories, household size, tenure in years, a supervisor status dummy, a permanent tenure status dummy, measured EGB, measured beta, measured delta, and measured financial literacy. We retain six significant factors and report the rotated factor loading matrix in Table 1.

While these estimated factors are nothing more than a low-dimensional summary of the variation in the data, examining the loadings allows for some meaningful interpretations. For example, the first factor loads primarily on fixed demographic characteristics such as gender and race (and conversely, these dimensions load primarily on this factor). The second retained factor loads primarily on age, length and type of tenure, and supervisory status. We thus interpret this as a composite measure of seniority. We find that the third retained factor loads on years of education, the measure of EGB, and the measure of financial literacy, and we interpret this as a composite measure of financial capability, measuring different aspects of financial sophistication. Finally, the fourth retained factor loads primarily on the estimated beta and beta × delta, and so we interpret it as a composite measure of time preference. The fifth factors loads on employee gender and household size, which we label household composition. Finally, the fifth factor aligns most with Hispanic background. We use these composite factors to consider heterogeneity in the treatment effect at a higher level of abstraction.

Our preanalysis plan was registered at the Social Science Registry AEARCTR-0002129. We prespecified that we would measure the heterogeneous effects of exponential-growth bias, time preferences, and financial literacy but we did not prespecify the factor analysis or the regressions using the factors. The reader should view these analyses as more exploratory.

3. Results

The design of our intervention allows us to investigate three questions. First, we examine whether selection into tool use varies by observable characteristics. Second, we measure the treatment effect among those who clicked the link in the email invitation to use the tool. Finally, we measure how the treatment effect varies with measures of a person's financial capability to determine whether the treatment is a substitute or complement to financial capability.

3.1. Selection into tool use

To examine selection into tool use, we regress tool use on individual characteristics using a logit specification and present our results in Table 2. First, we regress a binary variable that equals 1 for those who use the tool and zero otherwise on mean Alpha, mean Beta, and standardized financial literacy measure, and show these results in Column 1. None of the coefficients are statistically

Table 2
Selection into TOT sample.

	Logit		
	(1) Tool Participation	(2) Tool Participation	(3) Tool Participation
Tool Participation			
Mean Alpha	0.111	0.107	0.085
	(0.071)	(0.072)	(0.073)
Mean Beta	0.393	0.368	0.233
	(0.683)	(0.699)	(0.697)
Std. Financial Literacy	0.078	0.044	-0.009
	(0.056)	(0.061)	(0.063)
Age		-0.001	-0.009
		(0.006)	(0.006)
Male		-0.031	-0.059
		(0.121)	(0.125)
White		0.018	0.215
		(0.292)	(0.307)
Hispanic		-0.323	-0.171
		(0.390)	(0.408)
Black		-0.240	-0.015
		(0.312)	(0.325)
Some College or Associate		0.282	0.191
		(0.198)	(0.202)
Bachelor		0.240	0.008
		(0.168)	(0.177)
Post-Bachelor		0.186	-0.108
		(0.182)	(0.202)
Household Size		0.041	0.037
		(0.045)	(0.045)
Total Pay			0.003
			(0.003)
Tenure in Years			-0.006
			(0.009)
Team Leader			0.222
			(0.368)
Supervisor or Manager			0.415*
			(0.247)
Conditional - Tenure Group 2			0.577
			(0.494)
Permanent - Tenure Group 1			0.657
			(0.454)
Part-Time			0.845
			(0.882)
TSP Amount Pre-Rollout ($1,000/year)			0.048***
			(0.013)
Constant	0.252	0.096	-0.575
	(0.690)	(0.849)	(1.007)
Mean DV	0.667	0.668	0.668
Observations	1,435	1,393	1,392

Notes: Robust standard errors reported. Dependent variable in column heading. The omitted groups are: female, of other race, with high school education, holding non-supervisory position, and of other tenure group. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

significant, indicating no selection based on the primary variables that we hypothesized would play a role in insufficient retirement saving. We expand the regression to include age, gender, race, education, and household size (Column 2); these coefficients are statistically significant. In Column 3, we layer in employment attributes including total pay, tenure in years, leadership/manager, and tenure status. The only significant effect comes from preintervention TSP contribution amount. The effect is highly significant, with an additional standard deviation of TSP amount (SD = \$5,707.5) increasing the likelihood of using the tool by $e^{(5.7075\times 0.048)} - 1 = 32\%$. This finding indicates that those who are likely in greatest need of a course correction—those with low contributions—are less likely to use the tool.

3.2. Treatment-on-treated

We next estimate treatment-on-the-treated (TOT) effects, which represent the differences in contributions among the treatment group and the active control within the subsample of individuals who chose to interact with their version of the tool, rather than the intent-to-treat (ITT) effect among everyone invited to interact with the tool. In most experiments, the econometrician cannot observe which individual in the control group would take up treatment if offered the chance. Our active control design allows us to measure

Table 3
Average effects and heterogeneous effects by single dimensions of heterogeneity (TOT).

	TOT Main		TOT Heterogeneity				
	(1) Overall Sample	(2) Survey Sample	(3) Std. Alpha	(4) Std. Beta	(5) Std. Financial Literacy	(6) TSP Amount per year pre Rollout	(7) Bachelor or Higher
Post × Full Tool	174.184**	120.979	114.466	118.969	132.774	308.069*	-210.650
	(75.621)	(129.646)	(129.537)	(129.367)	(129.607)	(174.319)	(195.251)
Post × Attribute			-63.461	120.159	-166.267	0.073***	-179.543
			(84.566)	(108.571)	(102.292)	(0.018)	(201.044)
Post × Full Tool × Attribute			122.769	-152.713	328.038**	-0.022	496.098*
			(106.152)	(131.581)	(130.793)	(0.024)	(257.274)
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	7078.012	7577.489	7577.489	7577.489	7577.489	7577.489	7577.489
Permutation P Value	0.001	0.335					
FDR Sharpened Q-Value	0.081	0.259	0.248	0.248	0.081	0.259	0.1
R-squared	0.089	0.089	0.089	0.089	0.090	0.096	0.090
Observations	151,732	57,744	57,744	57,744	57,744	57,744	57,744

Notes: Robust standard errors in parentheses and clustered at the person level. The dependent variable is TSP amount. "Full" refers to the tool in the treatment condition. Col (1) reports the estimated TOT effects of all tool users, Col (2) of tool users who also answered the survey. Single-dimension attributes are collected from the survey. Col (3)–(7) report the heterogeneous TOT effects by attributes as specified in the corresponding column heading. All specifications include post dummy, year fixed effects, month fixed effects, and individual fixed effects. For Col (1) and Col (2), p-values from permutation inference of 1,000 times are reported. False Discovery Rate (FDR) Sharpened Q-values correct for table-wide multiple hypothesis testing, and include only the main statistical tests, not the controls. In Columns (1) and (2) they refer to the coefficient on Post × Full Tool and in Columns (3)-(7) they refer to the coefficients on Post × Full Tool × Attribute. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

this, creating a particularly strong TOT design and precise estimate. We focus on the TOT effect as it is better powered to detect differences between the conditions.

Using data at the individual-month level, we regress annualized TSP contributions on a post-intervention indicator, the treatment indicator, and the interaction between the two using a difference-in-difference framework.[9] The coefficient on the interaction term is our estimate of the treatment effect for the full treatment relative to the active control. We include year and month fixed effects to control for temporal variation in contributions, and individual fixed effects to control for between-person variation. We cluster standard errors at the person level.

Table 3 shows the main results of the treatment-on-treated analysis. The estimated effect of the treatment is a $174 increase in contributions per year ($p = 0.021$), which represents a 2.5 percent increase in annual contributions compared to the $7,078 average annual contribution and 0.2 percent of average annual pay (Column 1). We also report the mean of the dependent variable for the estimation sample and a p-value calculated using permutation inference at the bottom of the table. We do this by randomly relabeling participants as control and treatment 1,000 times and computing a counterfactual treatment effect for each simulation, which creates a distribution of treatment effects under the null hypothesis. Our estimated treatment effect exceeds all but the top 0.1 percent of our simulated treatment effects, giving rise to a p-value smaller than the estimate's p-value using asymptotic approximation. This effect is similar in magnitude to the effect found in Goda et al. (2014), who randomly assigned retirement income projections in the mail to University of Minnesota employees. Their treatment boosted average optional contributions by $85 annually, which was 3.6 percent of average optional contributions and 0.15 percent of pay. We also include False Discovery Rate (FDR) sharpened q-values as a row in the table, which correct for multiple hypothesis testing across our main statistical tests across the table (Anderson, 2008).[10]

Column 2 reproduces the same specification from Column 1 for the survey-response subsample, which is the sample we use to investigate heterogeneity in the treatment effect. While the TOT estimates in Columns 1 and 2 are similar, the estimate in Column 2 is a bit smaller in magnitude and the standard errors increase with the smaller sample, making the treatment effect no longer statistically significant at conventional levels.

3.3. Heterogeneity in treatment effects

Next, we estimate heterogeneous treatment effects (Columns 3–7). The coefficient of interest in each column is the three-way interaction (postintervention × treatment group × attribute), which may be interpreted as the increase in the treatment effect,

[9] The post-intervention indicator equals 1 for all individuals after the tool roll-out date. Thus, our analysis does not have staggered treatment timing and is not subject to concerns raised about two-way fixed effect models in difference-in-differences designs due to staggered treatment timing (Goodman-Bacon, 2021; Borusyak et al., 2022). Our estimates may be conservative given that some of the period after the tool rollout may be before an individual accesses the tool. To assess sensitivity to this issue, we estimate a model with staggered treatment timing across individuals based on the date of their first click through to the tool (Appendix Figure D.11). As expected, the effect estimated using Borusyak et al. (2022) is larger (Appendix Table D.14). However, the timing of first click is driven by individual differences between subjects and not random, so we do not prefer this specification.

[10] In Columns (1) and (2) the q-values refer to the coefficient on Post × Full Tool, and in Columns (3)-(7) the q-values refer to the coefficients on Post × Full Tool × Attribute.

relative to the active control, of a one-unit increase in the attribute. In Column 3, the attribute is measured exponential-growth bias; in Column 4, the attribute is the measured short-run discount rate, Beta; and in Column 5, the attribute is measured financial literacy. We standardize each of these attributes so a one-unit change corresponds to one standard deviation.

While we find no evidence of heterogeneous treatment effects with respect to measures of exponential-growth bias and present bias, we find evidence of a statistically significant difference in the treatment effect depending on one's level of measured financial literacy. The sign of the coefficient indicates that, rather than less financially literate employees benefiting from the increase in information, the treatment has a greater impact on more financially literate individuals, leading them to increase their contributions more. Specifically, having one standard deviation higher measured financial literacy increases the treatment effect by $328 in annual contributions. Because the treatment leads those with higher levels of measured financial literacy to make bigger changes in their contributions than those with lower measured financial literacy, the evidence suggests that the intervention complements rather than substitutes for financial capability.

The lack of significant heterogeneity by exponential-growth bias or present bias ran contrary to our expectation. The intervention was designed to help individuals accurately understand the mapping from retirement account contributions to retirement income. Exponential-growth bias distorts this understanding, tending to lead one to underestimate the future benefits of more-immediate sacrifices Goda et al. (2019). This may be a particularly acute issue for those with naive present bias as well.

In Column 6 we examine heterogeneity in the treatment effect based on preintervention contributions. We find no evidence of differences in the treatment effect between those who were contributing different amounts prior to the intervention (Column 6). We estimate heterogeneity by formal educational attainment and find that those with at least a bachelor's degree exhibit treatment effects that are $496 greater than those with lower levels of education, though the effect is only marginally significant (Column 7).[11]

Drawing on the latent factors described and estimated in Section 2.4.4, in Table 4, we include interaction terms for the three meaningful composite factors—seniority, financial capability, and time preference. As before, we include year and month fixed effects to control for temporal variation in contributions and individual fixed effects to control for between-person variation. The coefficients of interest are thus the triple-interactions of post × treatment × factor, which describes how the relative increase in the treatment group over the active control differs for those with a standard deviation higher level of the composite factor. In Columns 1 and 5, the estimated coefficient on the three-way interaction provides evidence that demographics is not associated with a statistically or economically significant heterogeneity in treatment effects. In Columns 2 and 5 the same can be said about seniority.[12] In contrast, one standard deviation higher in the financial capability factor is associated with a $412 stronger treatment effect. These results are consistent with those in Table 3, where measured financial literacy and education levels were associated with larger treatment effects. At this greater level of abstraction, we find that more financially capable employees benefit more from the increase in information; that is, the information intervention and financial capability are complements rather than substitutes. Third, we fail to find evidence that time preferences mediate the treatment effect. Finally, we include all four factors and their interactions with treatment and post simultaneously and find evidence that the only significant interaction is with financial capability and that this significant interaction is evident even when allowing heterogeneity on the other factors.[13]

4. Discussion

Our results are surprising in several ways. We find that selection into tool use favors those who save more, and who are therefore less likely to need a TSP saving correction. This finding goes against the overall efficacy of the tool in closing saving gaps as those who are at greatest risk of inadequate retirement savings are those who are least likely to use it. The treatment effect is increasing in measured financial literacy, education, and our composite financial capability metric, generated through exploratory factor analysis. We designed the intervention expecting that behavioral biases likely cause people to make suboptimal retirement-savings decisions, targeting EGB and procrastination, which have been shown in prior work to be associated with lower levels of retirement savings. We therefore hypothesized that an intervention designed to counteract those behavioral biases would improve decisions overall, but more specifically for those who were more biased. However, we found no evidence that either of these biases were correlated with the treatment effect.

Past literature has shown that measured financial literacy and financial capability is positively correlated with more retirement savings, while controlling for many other variables including income and age (Goda et al., 2019). However, complementarity between the treatment and financial capability implies that interventions like the one in this paper may be ineffective at helping employees who are most vulnerable. If this lower savings stems from uninformed decision-making, retirement saving outcomes are likely suboptimal, and so helping individuals who lack financial capability would be a natural policy goal. Our results suggest that simple online retirement-savings tools are not sufficient meet this goal despite their "just-in-time" design feature aimed (i.e., both versions of the tool included support for immediate implementation of contribution changes).

We speculate that a certain degree of financial capability is necessary to effectively digest the information contained in the treatment-version of the online tool. For those with higher measures of financial capability, providing information for how much

[11] Appendix Table A.5 replicates Table 3 but with the outcome in terms of standard deviations of TSP amount. Appendix Table A.8 replicates Table 3 but with the outcome in terms of TSP rate. Intent-to-treat versions of these tables are available in Appendix Table A.4, A.7, and A.10. The main effect is not significant in these analyses but the heterogeneous effects of financial literacy and education are similar in terms of sign and significance, as are the null effects of the other attributes.

[12] To aid interpretation, note that, in the control group, greater seniority is associated with a $294 smaller change after the experiment began versus before.

[13] Appendix Table A.6 replicates Table 4 but with the outcome in terms of standard deviations of TSP amount. Appendix Table A.9 replicates Table 4 but with the outcome in terms of TSP rate.

Table 4
Heterogeneous effects by factors (TOT).

	(1) TSP Amount ($/year)	(2) TSP Amount ($/year)	(3) TSP Amount ($/year)	(4) TSP Amount ($/year)	(5) TSP Amount ($/year)	(6) TSP Amount ($/year)	(7) TSP Amount ($/year)
Post × Full Tool	141.889	75.229	151.798	137.219	173.534	133.807	25.538
	(130.840)	(130.527)	(131.326)	(130.473)	(135.362)	(131.544)	(134.771)
Post × Demographics	-105.760						-107.469
	(95.464)						(96.001)
Post × Full Tool × Demographics	149.497						157.211
	(128.685)						(126.854)
Post × Seniority		-293.914***					-288.275***
		(99.988)					(99.769)
Post × Full Tool × Seniority		-38.885					-67.622
		(137.083)					(133.333)
Post × Financial Capability			-126.354				-113.895
			(97.740)				(96.591)
Post × Full Tool × Financial Capability			411.633***				364.711***
			(132.631)				(128.438)
Post × Time Preference				164.910			176.523
				(109.860)			(109.173)
Post × Full Tool × Time Preference				-180.815			-180.677
				(133.436)			(132.239)
Post × Household Composition					46.222		57.651
					(104.020)		(102.362)
Post × Full Tool × Household Composition					-101.637		-113.733
					(128.338)		(125.478)
Post × Hispanic Factor						-81.289	-78.221
						(93.459)	(84.823)
Post × Full Tool × Hispanic Factor						89.919	56.255
						(108.988)	(103.873)
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	7579.859	7579.859	7579.859	7579.859	7579.859	7579.859	7579.859
F-Statistic	1.350	0.080	9.632	1.836	0.627	0.681	
P-Value	0.246	0.777	0.002	0.176	0.429	0.410	
R-squared	0.089	0.094	0.093	0.092	0.092	0.092	0.107
Observations	56,131	56,131	56,131	56,131	56,131	56,131	56,131

Notes: Robust standard errors in parentheses and clustered at the person level. The dependent variable in the column heading. "Full" refers to the tool in the treatment condition. Factors are generated from the principal component analysis using single-dimensional attributes from survey as inputs. Factor loadings are reported in Table 1. All specifications include post dummy, year fixed effects, month fixed effects, and individual fixed effects. False Discovery Rate (FDR) Sharpened Q-values correct for table-wide multiple hypothesis testing, and include only the main statistical tests, not the controls. Q-values refer to the triple interaction terms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

annual income in retirement their current TSP contribution election provided was useful; in contrast, we found no evidence that this additional information was useful to those with lower measures of financial capability. Employees with lower financial capability may have been intimidated by the specific information pertaining to TSP election in the treatment version of the tool if self-knowledge about the saving vehicle was low Past research has shown that financial self-knowledge is low. Bhargava and Conell-Price (2021) find that 20–37 percent of nonparticipants in their employer's 401(k) program mistakenly believed that they had already enrolled. Furthermore, Bhargava et al. find that cosmetic user-interface design elements can have a large effect on employee savings rates, suggesting that many employees are making decisions in a haphazard or nondeliberative manner. To help employees with lower financial capability, online tools may require better automation whereby the fields in the online tool are autopopulated by the employee's administrative data. Such integration would lead to fewer steps, less reliance on financial language, and less need for employee self-knowledge. However, it is also possible that more expensive forms of intervention, such as one-on-one sessions and/or personalized materials, may be necessary to help those with lower financial capabilities.

5. Conclusion

We conducted a randomized controlled trial, inviting federal employees to use an online retirement-savings tool. Participants who received projections of their retirement income from their defined contribution plan saved $174 more annually than those who did not. Selection into the tool favored those who already had higher TSP contributions. The treatment effect was larger for those with higher measure of financial literacy and those who were more "Financial Capable," a factor generated by our factor analysis. This complementarity between the tool and financial capability indicators suggests that similar tools may be effective at helping the well-informed, educated, and financially literate to make optimal retirement-savings decisions, but unlikely to help those who are relatively uninformed, less educated, and less financially literate. Different approaches may be needed to help different populations.

One of the strengths of online tools is that they scale well: the marginal cost to the employer or plan manager is near zero. We find evidence of benefits for financially capable workers that may justify those costs. However, these findings suggest that more research and development regarding cost-effective ways to assist those with lower financial capability are needed to close gaps in retirement wealth.

Declaration of competing interest

None.

Data availability

The data that has been used is confidential.

Appendix. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jebo.2023.08.001.

References

Anderson, Michael L., 2008. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the Abecedarian, Perry Preschool, and early training projects. J. Am. Stat. Assoc. 103 (484), 1481–1495.

Austin, Percy, Arnott-Hill, Elizabeth, 2014. Financial literacy interventions: evaluating the impact and scope of financial literacy programs on savings, retirement, and investment. J. Soc. Polit. Econ. Stud. 39 (3), 290–314. Copyright - Copyright Council for Social and Economic Studies, Inc. Fall 2014; Document feature -; Last updated - 2022-11-21.

Bernheim, Douglas B., Garrett, Daniel M., 2003. The effects of financial education in the workplace: evidence from a survey of households. J. Public Econ. 87 (7), 1487–1519.

Bernheim, Douglas B., Garrett, Daniel M., Maki, Dean M., 2001. Education and saving:: the long-term effects of high school financial curriculum mandates. J. Public Econ. 80 (3), 435–465.

Beshears, John, Choi, James J., Laibson, David, Madrian, Brigitte C., 2009. The importance of default options for retirement savings outcomes: evidence from the United States. In: Social Security Policy in a Changing Environment. University of Chicago Press, Chicago, IL.

Beshears, John, Choi, James J., Laibson, David, Madrian, Brigitte C., 2013. Simplification and saving. J. Econ. Behav. Organ. 95, 130–145.

Beshears, John, Choi, James J., Laibson, David, Madrian, Brigitte C., Milkman, Katherine L., 2015. The effect of providing peer information on retirement savings decisions. J. Finance 70 (3), 1161–1201.

Bhargava, Saurabh, Conell-Price, Lynn, 2021. Serenity Now, Save Later? Evidence on Retirement Savings Puzzles from a 401(k) Field Experiment. Working Paper.

Bhargava, Saurabh, Conell-Price, Lynn, Mason, Richard T., Benartzi, Shlomo Save(d) by Design. Working Paper.

Borusyak, Kirill, Jaravel, Xavier, Spiess, Jann, 2022. Revisiting Event Study Designs: Robust and Efficient Estimation. Available at SSRN 2826228.

Brown, Jeffrey, Previtero, Alessandro, 2014. Procrastination, Present-Biased Preferences, and Financial Behaviors. Working Paper.

Bucher-Koenen, Tabea, Lusardi, Annamaria, Alessie, Rob J.M., van Rooij, Maarten, 2017. How financially literate are women? An overview and new insights. J. Consum. Aff. 51, 255–283.

Bundorf, Kate M., Polyakova, Maria, Tai-Seale, Ming, 2019. How do Humans Interact with Algorithms? Experimental Evidence from Health Insurance. Working Paper 25976. National Bureau of Economic Research.

Choi, James J., Laibson, David, Madrian, Brigitte C., 2006. Reducing the Complexity Costs of 401(k) Participation Through Quick Enrollment. NBER Working Paper 11979. National Bureau of Economic Research.

Choi, James J., Laibson, David, Madrian, Brigitte C., Metrick, Andrew, 2004. For better or for worse: default effect and 401(k) savings behavior. In: Wise, David A. (Ed.), Perspectives in the Economics of Aging. University of Chicago Press, Chicago, IL.

Choi, James J., Haisley, Emily, Kurkoski, Jennifer, Massey, Cade, 2017. Small cues change savings choices. J. Econ. Behav. Organ. 142 (C), 378–395.

Dolls, Mathias, Doerrenberg, Philipp, Peichl, Andreas, Stichnoth, Holger, 2018. Do retirement savings increase in response to information about retirement and expected pensions? J. Public Econ. 158, 168–179.

Duflo, Esther, Saez, Emmanuel, 2003. The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. Q. J. Econ. 118 (3), 815–842.

Falk, Armin, Becker, Anke, Dohmen, Thomas, Huffman, David, Sunde, Uwe, 2016. An Experimentally-Validated Survey Module of Economic Preferences. Working Paper.

Fernandes, Daniel, Lynch Jr, John G., Netemeyer, Richard G., 2014. Financial literacy, financial education, and downstream financial behaviors. Manag. Sci. 60 (8), 1861–1883.

Finkelstein, Amy, Notowidigdo, Matthew J., 2019. Take-up and targeting: experimental evidence from SNAP. Q. J. Econ. 134 (3), 1505–1556.

Fuentes, Olga, Lafortune, Jeanne, Riutort, Julio, Tessada, Jose, Villatoro, Felix, Forthcoming. Personalized information as a tool to improve pension savings: results from a randomized control trial in Chile. Econ. Dev. Cult. Change. https://doi.org/10.1086/720718.

Gale, William G., Levine, Ruth, 2011. Financial literacy: What works? How could it be more effective? How could it be more effective.

Goda, Shah Gopi, Flaherty Manchester, Colleen, Sojourner, Aaron, 2014. What will my account really be worth? Experimental evidence on how retirement income projections affect saving. J. Public Econ. 119, 80–92.

Goda, Gopi Shah, Levy, Matthew, Flaherty Manchester, Colleen, Sojourner, Aaron, Tasoff, Joshua, 2019. Predicting retirement savings using survey measures of exponential-growth bias and present bias. Econ. Inq. 57 (3), 1636–1658.

Goodman-Bacon, Andrew, 2021. Difference-in-differences with variation in treatment timing. J. Econom. 225 (2), 254–277.

Hackethal, Andreas, Haliassos, Michael, Jappelli, Tullio, 2012. Financial advisors: a case of babysitters? J. Bank. Finance 36 (2), 509–524.

Hastings, Justine S., Madrian, Brigitte C., Skimmyhorn, William L., 2013. Financial literacy, financial education, and economic outcomes. Annu. Rev. Econ. 5 (1), 347–373.

Jones, Damon, Molitor, David, Reif, Julian, 2019. What do workplace wellness programs do? Evidence from the Illinois workplace wellness study. Q. J. Econ. 134 (4), 1747–1791.

Levy, Matthew R., Tasoff, Joshua, 2016. Exponential growth bias and lifecycle consumption. J. Eur. Econ. Assoc. 14 (3).

Lusardi, Annamaria, 2008. Household Saving Behavior: the Role of Financial Literacy, Information, and Financial Education Programs. National Bureau of Economic Research. Working Paper 13824.

Lusardi, Annamaria, Mitchell, Olivia S., 2011a. Financial literacy and planning: Implications for retirement wellbeing. Technical Report. National Bureau of Economic Research.

Lusardi, Annamaria, Mitchell, Olivia S., 2011b. Planning and financial literacy: how do women fare? Am. Econ. Rev. 98 (2), 413–417.

Lusardi, Annamaria, Mitchell, Olivia S., 2014. The economic importance of financial literacy: theory and evidence. J. Econ. Lit. 52 (1), 5–44.

Madrian, Brigitte C., Shea, Dennis F., 2001. The power of suggestion: inertia in 401(k) participation and savings behavior. Q. J. Econ. 116 (4), 1149–1525.

Mandell, Lewis, Klein, Linda S., 2009. The impact of financial literacy education on subsequent financial behavior. J. Financ. Counseling Plan. 20 (1), 15–24+. Copyright - Copyright Association for Financial Counseling and Planning Education 2009; Last updated - 2023-03-01; SubjectsTermNotLitGenreText - United States–US.

Olckers, Matthew, 2021. On track for retirement? J. Econ. Behav. Organ. 190, 76–88.

OPM, 2015. Federal Employee Participation Patterns in the Thrift Savings Plan, 2008-2012.

Song, Changcheng, 2020. Financial illiteracy and pension contributions: a field experiment on compound interest in China. Rev. Financ. Stud. 33 (2), 916–949.

Stango, Victor, Zinman, Jonathan, 2009. Exponential growth bias and household finance. J. Finance 64 (6), 2807–2849.

Thaler, Richard, Benartzi, Shlomo, 2004. Save more tomorrow: using behavioral economics to increase employee saving. J. Polit. Econ. 112 (1), S164–S187.

Willis, Lauren E., 2021. Alternatives to financial education. Loyola Law School, Los Angeles Legal Studies Research Paper (2021-11).