GFLEC

GLOBAL FINANCIAL LITERACY EXCELLENCE CENTER

THE EFFECT OF FINANCIAL EDUCATION ON THE QUALITY OF DECISION MAKING

Sandro Ambuehl, B. Douglas Bernheim, and Annamaria Lusardi

WP 2016-2

August 1, 2016

GW

GFLEC Working Paper Series

The Effect of Financial Education on the Quality of Decision Making

Sandro Ambuehl

B. Douglas Bernheim

Annamaria Lusardi*

August 1, 2016

Abstract

We introduce a method for measuring the quality of financial decision making built around a notion of *financial competence*, which gauges the alignment between individuals' choices and those they would make if they properly understood their opportunities. We use it to document the potential pitfalls of the types of brief rhetoric-laden interventions commonly used for adult financial education. Motivational rhetoric can render the effects of such interventions indiscriminate even when people appear to understand and internalize the targeted concepts. Conventional methods of evaluation involving financial literacy, self-reported decision strategies, and directional effects on choices do not reliably detect these deficiencies.

JEL Codes: C91, D03, D04, D14, D60, D91, I21

*Ambuehl: Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305, sambuehl@stanford.edu. Bernheim: Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305, bernheim@stanford.edu. Lusardi: The George Washington University School of Business, 2201 G Street, NW, Suite 450E, Washington, DC 20052, alusardi@email.gwu.edu. We thank Charles Sprenger, Steven Sheffrin, Glen Weyl, as well as participants at the Research Forum on the Effectiveness of Financial Education at the University of Arizona, the Stanford Institute for Theoretical Economics, the Journées Louis-André Gérard-Varet in Aix-en-Provence, the New York University, the Murphy Institute's conference on Expanding the Frontiers in Behavioral Public Economics, the Cherry Blossom Financial Education Institute at the George Washington University, the Roybal Conference on Complexity in Decision Making at the University of Southern California, the Stanford Institute for Economic Policy Research, the TIAA Institute Fellows Symposium, the 2016 ASSA Meetings in San Francisco, the George Washington University Financial Literacy Seminar Series, and the Workshop on the Behavioral Economics of Financial Markets at the University of Zurich for helpful comments and suggestions. Fulya Yuksel Ersoy provided excellent research assistance. This work was funded by the Department of Economics at Stanford University, and was conducted under the Stanford IRB protocol 29615.

"A little learning is a dangerous thing; Drink deep, or taste not the Pierian spring: There shallow draughts intoxicate the brain, And drinking largely sobers us again"

– Alexander Pope, *An Essay on Criticism*, (1709)

1 Introduction

Low levels of financial literacy in the United States and the rest of the world raise doubts about the general quality of financial decision making. Financial education aims to improve decisions by helping consumers acquire the basic knowledge and skills they need to understand the choices they face. A large and growing literature finds mixed evidence that financial education interventions affect behavior (Hastings, Madrian and Skimmyhorn (2013), Lusardi and Mitchell (2014) provide reviews). Discussions of their welfare effects are typically informal and often colored by paternalistic judgments and preconceptions – for example, that people are better off with high saving and balanced portfolios, or that a better understanding of financial concepts necessarily promotes better decisions. Yet it is also possible that particular interventions alter behavior through mechanisms that involve indoctrination, exhortation, deference to authority, social pressure, or psychological anchors. If so, their benefits are unclear.

These concerns are particularly acute for workplace interventions, which provide the lion's share of adult financial education in the U.S.[1] Employers effectively treat *brevity* as a design constraint: thorough educational programs are not only costly but also time-consuming, which makes them unappealing to workers.[2] To compensate for brevity, these programs generally focus on simple heuristics accompanied by highly motivating messages. The intent is to make the substantive material engaging, memorable, and actionable. Yet compelling rhetoric may also distract from substance and promote a one-size-fits-all response, which may be excessive for some and even directionally inappropriate for others.

This paper makes two main contributions. First, we introduce a new method for measuring the quality of financial decision making. This contribution is important because rigorous analyses of decision-making quality are missing from most studies of financial education, likely due to the

[1] In a 2013 survey of 407 retirement plan sponsors covering more than 10 million workers by Aon Hewitt, 77% of providers offered on-site financial education seminars or meetings (Austin and Evens, 2013). In the 2015 FINRA National Financial Capability Study, 40.24% of respondents aged 20 - 65 who have received financial education did so through an employer.

[2] A meta-analysis by Fernandes, Lynch and Netemeyer (2014) finds that the average financial education program involves only 9.7 hours of instruction. That time is usually divided among a long list of complex topics. For example, Skimmyhorn (2015) reports that a financial education program used by the U.S. military covers compound interest, the focus of our current study, along with a collection of several more complex topics – retirement concepts, the Thrift Savings Plan, military retirement programs, and investments – all within a single two-hour session.

limitations of existing methods, as discussed in Section 2.3. The essence of our approach is to assess a consumer's willingness to pay (WTP) for equivalent consumption opportunities. We package these opportunities as assets and design them so that a knowledge of targeted financial principles is required to understand that one is a simplified version of the other. Someone who both possesses and fully operationalizes that knowledge will consistently ascribe exactly the same value to the equivalent simply and complexly framed opportunities regardless of their preferences. When these WTPs differ systematically, the magnitude of the divergence provides a measure of *financial competence* with respect to the targeted principles: it indicates the extent to which a consumer's incomplete operational command of those principles exposes her to decision error.

As we emphasize, our approach has several virtues. First, it is non-paternalistic. The types of external judgments of consumers' choices that are common in policy discussions, such as whether they are "sufficiently patient" or "save enough," are entirely avoided. Second, it imposes modest information requirements. By comparing a consumer's choices for equivalent tasks, we avoid the need for parametric models of decision making. Third, it is simple, intuitive, and easily implemented. Fourth, it yields a quantitative measure of financial competence that one can formally interpret as an index of consumer welfare with the framework of Bernheim and Rangel (2009) (see also Bernheim, 2009 and Bernheim, 2016).[3] As we explain in the next section, our methods also offers important advantages over existing approaches to measuring the quality of financial decision making, including the examination of dominated choices (Ernst et al., 2004; Calvet et al., 2007, 2009; Agarwal et al., 2009; Baltussen and Post, 2011; Choi et al., 2011), the evaluation of WARP-consistency (Choi et al., 2014), and structural modeling (Song, 2015).

Our second main contribution is to document, through an experiment, the potential pitfalls of the types of brief rhetoric-laden interventions that are commonly used for adult financial education, and to demonstrate that conventional methods of evaluation may fail to detect their deficiencies. Our experiment involves an educational intervention that focuses on compound interest, one of the fundamental concepts in personal finance. It resembles typical employer-sponsored interventions with respect to its brevity, as well as its emphasis on heuristics and motivational messages. It also appears to be highly effective according to conventional outcome measures: treated subjects perform substantially better on an incentivized financial literacy test, they report applying their newly gained knowledge when performing the decision tasks we assign them, and their average WTPs for interest-bearing assets change in a direction that counteracts the previously documented tendency to underestimate compounding, a phenomenon known as *exponential growth bias* (Wagenaar and Sagaria, 1975; Eisenstein and Hoch,

[3] A handful of other studies undertake comparisons between simply and complexly framed choices (Hastings and Tejeda-Ashton, 2008; Bertrand and Morse, 2011; Abeler and Jaeger, 2014; Kalaci and Serra-Garcia, 2016), but none uses equivalent valuation tasks to infer the welfare losses resulting from complex framing, or the effect of educational intervention on those losses.

2007; Stango and Zinman, 2009; Almenberg and Gerdes, 2012; Levy and Tasoff, 2016). Nevertheless, using our approach, we find that the intervention does not, on average, improve the quality of decision making.

A possible explanation for this finding is that subjects may interpret motivational rhetoric as substantive advice and, even when their tested knowledge improves, emerge with an insufficient *operational* understanding of financial concepts to make appropriate adjustments. To explore this hypothesis, we implement two additional variants of the intervention, one that retains its substantive elements but omits the motivational rhetoric, and another that retains the motivational rhetoric but omits almost all of the substance. We show that the effects on financial literacy and self-reported decision strategies are primarily attributable to the substantive elements of instruction, as one would hope. However, in sharp contrast, the effects on financial choices are primarily attributable to the non-substantive elements. In particular, the intervention's motivational rhetoric increases subjects' WTP for interest-bearing assets regardless of the extent to which any particular individual initially understates or overstates the effects of compounding.[4] This indiscriminate response is beneficial in some cases and harmful in others; on average, there is no benefit.[5] When stripped of motivational rhetoric, exclusively substantive instruction has some effect on behavior, and it does reduce reliance on simple interest calculations (the most common type of mistake), but it fails to promote reliance on correct compound interest calculations, instead increasing the prevalence of other mistakes. As a result, its impact on WTPs for interest-bearing assets is directionally haphazard and, on average, welfare-neutral.

Thus, while financial literacy undoubtedly plays an important role in decision making (as shown by Lusardi and Mitchell, 2011), the associated mechanisms are complex and mediated by a variety of other factors. Educational interventions that achieve similar improvements in tested comprehension may have dissimilar effects on behavior, depending on the particular manner in which each intervention motivates participants, and whether it helps them learn to internalize and operationalize conceptual knowledge rather than directional imperatives. Accordingly, one would expect to find sharp differences between the effects of adult financial education programs and high school courses: as we have noted, the former typically compensate for brevity with simple heuristics and motivational rhetoric; in contrast, the latter often span a full semester, permitting a more expansive and in-depth treatment of subject matter, as well as more effective pedagogy, including practice and discussion. While the literature studies these two settings separately (beginning with Bernheim, Garrett and Maki, 2001, and

[4]As in Goda et al. (2015) and Levy and Tasoff (2015), we document considerable heterogeneity with respect to the perceived benefits of compounding.

[5]Song (2015) also offers evidence that an educational intervention of involving compound interest has an indiscriminate impact: the effect on measured saving is not closely related to the gap between actual and optimal rates implied by a parameterized life-cycle consumption model.

Bernheim and Garrett, 2003), it has only recently begun to explore the heterogeneity of approaches within each category, and to examine how the effects of an intervention depend on its design and constituent components. Consistent with our findings, recent work by Brown et al. (2014) shows that the effects of high school financial education on behavior are most pronounced when schools offer full courses taught by trained teachers. More generally, the considerations highlighted in the current study may help to explain why different authors reach different conclusions about the effects of financial education when studying different programs; see in particular Duflo and Saez (2003), Bayer, Bernheim and Scholz (2009), Goda, Manchester and Sojourner (2014), Cole and Shastry (2012), Cole, Sampson and Zia (2011), Skimmyhorn (2012), Servon and Kaestner (2008), Collins (2010), Lührmann, Serra-Garcia and Winter (2014), Mandell (2009), Drexler, Fischer and Schoar (2014), Carlin, Jiang and Spiller (2014), Heinberg et al. (2014), Lusardi et al. (forthcoming), and Bertrand and Morse (2011).

The remainder of the paper is organized as follows. Section 2 precisely defines the concept of financial competence, discusses its measurement, and explains its formal connection to consumer welfare. It also compares our approach to other methods for assessing the quality of financial decision making. Sections 3 describes the design of our experiment and section 4 discusses its implementation. Sections 5 analyzes the effects of the treatments on standard outcomes measures, including test scores, self-reported decision strategies, and average choices. Section 6 examines effects on the quality of financial decision making. We address the important issue of generalizability at some length in Section 7. Section 8 discusses the implications of our research and concludes.

2 The Definition and Measurement of Financial Competence

Our first objective is to devise a general framework for assessing the quality of financial decision making. We seek to formalize the intuitive notion that a good decision maker is one who avoids mistakes. This objective requires us to depart from classical consumer theory: if all choices reveal preferences, then none are mistaken, and any apparent inconsistencies must reflect our own misconceptions about the consumer's aims. As a general matter, we will say that a consumer displays financial competence with respect to targeted financial principles if they make equivalent choices from equivalent opportunity sets in contexts where the targeted principles govern the equivalence. In this section, we begin by explaining the basic idea behind our approach, and then address important complicating considerations. We conclude with a comparison between our approach and others that appear in the literature.

2.1 The Basic Idea

In textbook treatments of decision making, consumers select from menus of consumption bundles consisting of the goods they desire intrinsically. Actual financial decision making is less direct. Instead of selecting consumption bundles, consumers choose *consumption instruments*. These are derivative goods such as financial assets that provide the means to secure consumption bundles. The distinction between choosing a consumption bundle versus a consumption instrument is of no consequence for someone who fully appreciates the relationships between choices and outcomes. However, in the context of financial decisions, those relationships are governed by principles that many people demonstrably do not understand (see, for example, Lusardi and Mitchell, 2011). We are concerned here with detecting and evaluating mistakes emanating from such misunderstandings.

The essence of our approach is to assess a consumer's willingness to pay (WTP) for pairs of equivalent consumption opportunities, where one member of each pair is a *simplified* version of the other. We say that the comparatively complex version involves complex framing, and that the comparatively simple version involves simple framing. To be clear, our notion of equivalence is predicated on the assumption that people care intrinsically about the goods they purchase with the income they receive, but not about the packaging of that income. We describe the consumption opportunities as income-generating assets, but it is worth emphasizing that valuation is not simply a matter of consulting a market price or solving a math problem. Rather, for our subjects, it involves an expression of preference concerning factors such as timing and the perceived reliability of the promised claims.

We design the assets so that a knowledge of targeted financial principles is required to understand why each pair is equivalent. Someone who both possesses and fully operationalizes that knowledge will consistently ascribe the same value to both assets regardless of their preferences. When WTPs for equivalent assets differ systematically, the magnitude of the divergence provides an intuitively appealing measure of the individual's competence to make good decisions in contexts involving the pertinent principles. Moreover, as we explain below, the divergence has a precise welfare interpretation: it indicates the extent to which the consumer's incomplete operational command of those principles exposes her to losses.

To illustrate, say we are concerned that people poorly understand the concept of compound interest, and that this limitation causes them to make suboptimal investment decisions. To evaluate this possibility, we might assess the consumer's monetary equivalents for pairs of consumption opportunities such as the following: asset C represents a \$10 investment that promises a return of 6% per day compounded daily for 15 days, while asset S simply promises \$24 in 15 days. Ordinarily, a consumer will be willing to choose each asset over a fixed sum of money if and only if the sum does not exceed some threshold value, call it p^* for the first asset and q^* for the second. A quick calculation reveals that the two assets are equivalent, subject to rounding. Thus, swapping out one for the other in a decision problem changes framing while leaving opportunities intact. We would therefore hope to find $p^* = q^*$. As a general matter, any educational intervention that successfully provides subjects with an operational understanding of compound interest should bring p^* and q^* into closer alignment.

How should we interpret a divergence between p^* and q^*? Consumers who demonstrably misunderstand compound interest are presumably susceptible to error when making decisions involving asset C. What about asset S? Assume for the moment that all income is consumed when it is received. (We acknowledge that this assumption is unrealistic for some consumers, and discuss the implications of relaxing it in the next subsection.) In that case, as long as the income stream flowing from each alternative is clearly spelled out (as it is for asset S), evidence of the same misunderstanding gives us no reason to second-guess choices. We can therefore assess the magnitudes of the errors consumers make when evaluating asset C by treating the monetary equivalent for asset S as the correct value.

Notice in particular that $\left|p^* - q^*\right|$ represents the largest possible welfare loss the consumer may suffer when choosing between receiving an amount of money $\$d$ on the one hand and asset C on the other. To see this point, suppose first that $q^* < p^*$. If $d \leq q^*$ or $d \geq p^*$, there is no welfare loss, because the consumer would make the same choice regardless of which asset he considers.[6] Mistakes occur when $q^* < d < p^*$. In this case, the consumer chooses asset C over $\$d$, even though she would willingly exchange the returns to asset C for $\$d$ if she fully appreciated the consequences of her choice. If she started out with her best option, $\$d$, she would be willing to give up $\$\big(d - q^*\big)$ to avoid swapping that cash for the income stream that both assets promise. Hence, $\$\big(d - q^*\big)$ is the equivalent variation associated with the swap: it measures the dollar loss the consumer regards as equivalent to suffering the consequences of decision error. The loss is greatest when $d = p^*$.

Next suppose that $p^* < q^*$. Reasoning as above, we see that mistakes occur only when $p^* < d < q^*$. In this case, the consumer chooses $\$d$ over asset C, even though she would willingly exchange $\$d$ for the returns to asset C if she fully understood them. If she started out with those returns, she would require $\$\big(q^* - d\big)$ as compensation for switching to $\$d$. Assuming income effects are negligible over the relevant range, compensating and equivalent variation coincide; $\$\big(q^* - d\big)$ then measures the dollar loss the consumer regards as equivalent to suffering the consequences of decision error.[7] The loss is again greatest when $d = p^*$.[8]

[6] Technically, in the special case where $d = p^*$, she would definitely choose asset S over $\$d$, and is willing to choose asset C over $\$d$.

[7] In other words, we assume that a consumer who is indifferent between C and $\$(d+r)$ immediately is also indifferent between a bundle consisting of C with a loss of $\$r$ immediately, and $\$d$ immediately (because the immediate income for both options is reduced by the same amount, $\$r$), which implies that $\$r$ is the equivalent variation associated with the switch from C to $\$d$ immediately.

[8] For the purpose of the application considered in this paper, the assumption of negligible income effects is reasonable. More generally, one can handle the case of non-negligible income effects by adjusting our valuation-elicitation procedure.

Of course, the largest possible welfare loss generally overstates the actual loss. Another possibility is to compute the consumer's average or expected loss. Naturally, the expected loss depends on the process generating the consumer's opportunities. In the context of our experiment, the value of d is drawn from a uniform distribution. The probability of incurring a loss is therefore proportional to $\left|p^* - q^*\right|$, and the expected loss conditional upon suffering one is $\left|p^* - q^*\right|/2$; thus, the expected loss is proportional to $\left(p^* - q^*\right)^2$. More generally, one can think of $\pi\left(p^* - q^*\right)^2$ (where π is the density of the CDF governing the distribution of d at $d = q^*$) as a second-order approximation of the expected welfare loss, much in the spirit of Harberger's (1964) well-known formula for the deadweight loss of a commodity tax.[9]

Rigorous foundations for the preceding approach to evaluating the quality of financial decision making are found in Bernheim and Rangel (2009) and Bernheim (2016) (see also Bernheim, 2009, and Bernheim and Rangel, 2004). Within that framework, one classifies a decision as a mistake when it has two distinctive features. First, it is predicated on a characterization of the available options and the outcomes they imply that is inconsistent with the decision maker's information ("characterization failure").[10] In our experiment, we establish the existence of characterization failure for decisions involving asset C by designing the asset so that consumers require a working knowledge of compound interest to understand its implications, and then demonstrating that they lack that knowledge. Such failures raise the *possibility* that a mistake may have occurred, but does not guarantee it, because one can make the right decision for the wrong reason. Thus the definition of a mistake includes a second feature: there is some other option in the opportunity set that the decision maker would select over the mistakenly chosen one in settings where characterization failure does *not* occur. In any given application, choice problems without demonstrated characterization failure constitute the "welfare-relevant domain." Under our stated assumptions (which we revisit below), decisions involving asset S fall within this domain, because we have designed the asset so that its implications for consumption are transparent, even without a knowledge of compound interest. When applying the framework, one constructs a welfare criterion based on the welfare-relevant choices.[11] One potential use of that criterion is to quantify the welfare loss the consumer incurs when making mistakes, just as we have done in proposing measures of financial competence. The methodological contribution of the current

[9] Fix the value of q^*, and assume that the CDF governing the distribution of d is twice differentiable at $d = q^*$. Taking a second-order Taylor expansion of the expected welfare loss as a function of p^* in a neighborhood of q^*, we obtain $\frac{\pi(q^*)}{2}\left(p^* - q^*\right)^2$.

[10] This term first appears in Bernheim (2009), but the concept is present in Bernheim and Rangel (2004) and Bernheim and Rangel (2009).

[11] The criterion is known as the "unambiguous choice relation." If welfare-relevant choices are internally consistent in the sense of WARP (Arrow, 1959), the approach amounts to revealed preference on a limited domain. The approach also allows for welfare-relevant choices that are internally inconsistent, in which case the resulting criterion admits a degree of ambiguity.

paper is to devise a simple method of applying the Bernheim-Rangel framework, involving paired valuation tasks, that permits one to make direct inferences about welfare without modeling behavior.

2.2 Complications

For the purpose of the preceding discussion, we assumed that people consume income when they receive it. Violations of this assumption raise the possibility that misunderstandings of opportunities might extend even to simply framed options such as asset S, or that people might have mistaken expectations about their own future behavior. Either possibility calls into question the premise that someone who misunderstands asset C would nevertheless comprehend the consequences of purchasing asset S. Even so, given the objective equivalence between assets C and S, the difference between their valuations continues to provide an intuitively appealing measure of financial competence. Moreover, as we explain next, the welfare interpretation of this measure is reasonably robust.

Weaker assumptions While we plainly cannot dispense with assumptions entirely, we will proceed here based on a weaker premise. Specifically, we will assume that people evaluate investment options in two steps: first, they form beliefs about the cash flows an investment will generate; second, they think about how they would spend that income. Taking into account their expectations about spending, and assuming away any *persistent framing effects* (which would arise if the framing of the original asset continued to influence subsequent choices even after cash flows were realized), we can then model the consumer as having reduced-form preferences over cash flows. We can write those preferences as $U(y)$, where y is a vector of the time-dated and state-contingent cash flows the consumer expects to receive. In effect, our method amounts to performing welfare analysis based on the as-if reduced-form utility function U implied by choices made in the simplified decision problems.

Stability of U In principle, financial education could change the anticipated uses of current and/or future cash, altering U. That state of affairs would render the welfare effects of financial education more complex and difficult to measure. Fortunately, the stability of U is a testable property. As discussed in subsequent sections, we find no evidence that financial education meaningfully influences either WTPs for simplified decision problems, or the amount of time subjects take to arrive at these valuations. Moreover, few subjects report deploying the targeted principles when evaluating simply framed options. These findings are consistent with the hypothesis that our subjects have stable expectations about the disposition of future income. For example, they may expect to consume the vast bulk of incremental income when they receive it, as assumed in the previous subsection.

Mistaken expectations Unfortunately, there is no guarantee that U embodies *correct* expectations concerning the deployment of income. One can offer theoretical arguments both for and against this possibility. On the one hand, consumers with limited financial competence may not understand how incremental cash translates into incremental consumption, except in the special case where they spend it all immediately. On the other hand, a consumer may learn her marginal utility of income from experience to a reasonable degree of accuracy, even without appreciating the underlying details. Unlike stability, this is an issue we cannot resolve empirically with the data at hand.

The possibility that U might embody *incorrect* expectations about the deployment of income potentially raises conceptual difficulties. Within the Bernheim-Rangel framework, it implies that characterization failure may be present in both simply and complexly framed decision problems. Yet even then, our measure of financial competence has a meaningful welfare interpretation: it indicates the (maximal or expected) welfare loss the consumer would experience if her understanding of the relationship between intermediate and final outcomes (that is, between income flows and ultimate consumption) were correct. While that is not necessarily an accurate measure of the overall welfare loss, it does tell us how the consumer's misunderstanding of a complexly framed investment contributes to the overall loss. To understand why this is valuable, suppose our object is to evaluate a policy initiative that seeks to improve financial choices through a multi-faceted intervention, the individual components of which focus on different different aspects of the problem. In that context, it is useful to know how each component performs individually in addressing its objective, without regard to whether it inadvertently mitigates or magnifies the failings of another component. By treating U as, in effect, "true utility" (even if it is not), our measure of financial competence provides that information.

Other sources of decision error A somewhat different concern is that U might reflect behavior the analyst does not deem welfare-relevant. For example, if consumers are time-inconsistent, and if one adopts the view that choices with immediate payoffs are susceptible to "present bias," then the instantaneous WTP for an asset with future payoffs understates its "true value" to the consumer. In that case, an intervention that increases p can arguably enhance welfare even if it increases $\left|p^* - q^*\right|$ (recall that p and q are, respectively, the WTPs for assets C and S). Notice, however, that a conceptual problem arises in this example only because it misapplies our method. One always needs to structure the diagnostic decision tasks so that the simply framed versions are deemed welfare-relevant. Given the example's assumptions, one would accomplish this objective by specifying that "early" cash flows are received after a delay of, say, one to two days rather than instantaneously, so that measured welfare losses are consistent with the *long-run criterion* (O'Donoghue and Rabin, 1999). That is precisely the approach we take in the current experiment.[12]

Corroboration of assumptions All of the preceding hinges on the premise that people evaluate complexly framed investment options in the manner described at the outset of this section. That premise is consistent with several findings presented in subsequent sections. First, subjects take much longer to perform valuation tasks when they involve complexly framed assets. Second, while education targeting the principles that govern the relationship between simply and complexly framed assets does not alter decision times for the former, it does for the latter. Third, subjects report using those principles with much higher frequency when evaluating complexly framed rather than simply framed assets. All of these findings are consistent with the view that subjects evaluate their WTPs for complexly framed assets by translating them into simply framing just as we have assumed, and they are difficult to reconcile with specific alternative hypotheses.[13]

Costs of cognition One potential limitation of our method is that it does not encompass the cognitive costs of contemplation, which education likely influences. However, as long as one is willing to assume that cognitive effort enters preferences separably, our approach still measures an important component of welfare. Though methods of measuring cognitive costs are still in the infancy (see, for example, Caplin and Martin, 2015, and Caplin and Dean, 2015), one could in principle assess them separately and factor them into a more complete analysis of welfare. With respect to our current experiment, we doubt the adjustment would be significant. While we cannot measure the costs of contemplation directly, we can (as noted above) track the speed with which subjects make their choices. The range of variation in the average decision time across treatments is on the order of 20 seconds. For subjects who are accustomed to working for five dollars per hour, the value of this incremental contemplation is a matter of pennies.

[12] As noted in Bernheim (2016), the philosophical case for applying the long-run criterion is far from clear-cut (see also Bernheim and Rangel, 2009, and Bernheim, 2009). Ideally, an analyst who takes a different stand concerning the welfare-relevant domain would restructure the diagnostic tasks accordingly. One who remains agnostic as to whether time inconsistency involves characterization failure would include two versions of each simply framed task, one in which "early" cash flows are instantaneous, and another in which they are delayed. The welfare criterion would then admit a degree of ambiguity, corresponding to the gap between the two valuations. For a related application of Bernheim and Rangel's *unambiguous choice criterion*, see Online Appendix C.4.

[13] As an example, consider the following alternative: subjects translate simply framed assets into complex framing, and then evaluate the latter based on heuristics. This possibility is not completely far-fetched, in that complex framing may be more familiar than simple framing. However, it is difficult to reconcile with the three findings reported in the text. Significantly, even if it were valid, it would not overturn the welfare interpretation of our financial competence measures. Notice in particular that $\left|p^* - q^*\right|$ and $\left(p^* - q^*\right)^2$ are symmetric in p and q. Thus, we arrive at the same welfare loss regardless of whether we think poor understanding of financial principles distorts the translation of complexly framed options into simply framed ones, or vice versa; we simply have a different view as to which choices are suboptimal.

2.3 Comparisons to Other Approaches

Economists have developed and deployed different methods of evaluating the quality of financial decision making for different purposes. With respect to the current application, our approach offers important advantages.

The most common alternative is to evaluate the prevalence of dominated choices; see Ernst et al. (2004), Calvet et al. (2007, 2009), Agarwal et al. (2009), Baltussen and Post (2011), Choi et al. (2011), and Aufenanger et al. (2016). The essence of this approach is to select diagnostic tasks that remove personal preferences from the mix. In effect, each decision boils down to solving a math problem that has one and only one correct answer. Consequently, the approach amounts to administering an *incentivized* test of financial literacy. Conversely, every incentivized financial literacy test, including the one we administer as part of this experiment, consists of decision tasks in which a single choice – the correct answer – is the dominant option.

In contrast, the vast majority of real-world financial decisions are not simply math problems: the 'right' choice almost always depends on preferences. Thus, a central issue when evaluating financial education interventions is whether people operationalize pertinent knowledge and concepts when preferences remain in the mix. They may not.[14] Posing a problem that has no objectively correct answer may reduce the resemblance to textbook examples, making the applicable principles harder to recognize. People may be less likely to deploy mathematical tools when mathematics potentially govern only one amongst several aspects of evaluation. Consideration of preferences may also activate specialized heuristics or psychological mechanisms, such as motivated reasoning (Kunda, 1990), that sweep relevant principles into the background, even if they are invoked. An important advantage of our approach is that, unlike the dominance agenda, it permits us to evaluate the quality of decision making rigorously even when preferences remain in the mix.

In principle, by deploying structural methods involving explicit models of preferences and choices, one could achieve the same advantage. Unfortunately, in any given application, that approach may necessitate much stronger assumptions than many analysts are willing to make or accept. To our knowledge, Song (2015) is the only existing empirical study that employs this approach in the context of financial education. He uses a life-cycle consumption model to evaluate the welfare effects of changes in retirement contributions resulting from an educational intervention targeting compound interest. His analysis hinges on the accuracy with which a particular life-cycle model, calibrated with data drawn from other choice domains, describes lifetime opportunities, unobserved future choices,

[14] This disconnect has been observed in other contexts. Enke and Zimmermann (2013) show that many people tend to neglect correlations even in simple settings, despite knowing how to account for them. Taubinsky and Rees-Jones (2015) find that many consumers underreact to excise taxes, even though they can properly compute tax-inclusive prices. Likewise, in the current context, consider the contrast between the conclusions we reach when evaluating our financial education intervention based on an incentivized test of financial literacy (effectively choice problems with dominant options) and our measures of financial competence.

and 'true' preferences.[15] By focusing on consistency within paired valuation tasks, our approach avoids the need to endorse a particular structural model and allows us to proceed under much weaker assumptions.

There are, of course, other notions of internal consistency such as WARP and GARP, and these have also been used to assess the quality of financial decision making (Choi, Kariv, Mueller and Silverman, 2014). These tools complement our approach because they allow one to examine the consistency of non-equivalent choices made in a fixed decision frame, rather than the consistency of equivalent choices across different frames. However, for the following reasons, measures of within-frame consistency are less well-suited to the task of assessing financial education interventions than our approach. First, they are not designed to detect the types of decision making failures that primarily concern us. A consumer who misunderstands a financial concept in a consistent manner will nevertheless respect such axioms. For example, one who incorrectly believes that bundle i will ultimately lead to a better consumption bundle than bundle j, perhaps because she uses the simple interest formula to assess compound interest, will choose i over j, and will never choose j when i is available; therefore, her choices among bundles will satisfy WARP. Second, financial education does not target conformance with WARP directly, and non-conformance may result from a variety of considerations that are unrelated to the consumer's understanding of specific financial principles (such as incompleteness of underlying preferences). In contrast, our approach allows one to design the paired valuation tasks so that the targeted principles govern their equivalence. Third, our approach more readily yields measures of non-conformance that are interpretable as welfare losses.[16]

One important point of differentiation among the various studies mentioned above is that some evaluate real-world decisions while others examine choices in experimental tasks. The current application of our method falls into the latter category.[17] While the use of experimental data permits us to proceed with fewer assumptions and facilitates sharper conclusions, there is also a cost, in that it is less obvious whether the conclusions generalize to the choices that actually matter. We will defer our discussion of this concern to Section 7, which broadly addresses questions of generalizability.

[15] He concludes that the intervention improved welfare on average even though its effect on behavior was indiscriminate. Actual changes in saving were not closely related to the optimal changes prescribed by the life cycle model, and the education intervention induced some subjects to oversave. Another interpretation is that the life-cycle model poorly captured actual objectives.

[16] To be clear, some measures of non-conformance with GARP, such as the Afriat (1972) critical cost efficiency index, do have efficiency interpretations; see, e.g., Choi et al. (2014) for a related application. Moreover, Echenique et al. (2011) provide a measure of non-conformance that is interpretable as the maximal amount of money one can extract from a decision maker with specific violations of GARP.

[17] One can also use our method to assess real-world choices, but the implementation is more challenging. Admittedly, it may be easier to find naturally occurring opportunities to study the frequency of dominated choices. That said, dominance is typically hard to establish in the field, because the complexity of the real world invites many possible rationalizations for ostensibly poor choices. As an example, consider the use of payday loans by consumers with unused credit card balances. While agreeing that this practice is generally ill-advised, we question whether one can legitimately categorize it as dominated, as some have claimed (see Ernst et al., 2004). In principle, it could be rational for a consumer to preserve some of the instant liquidity credit cards offer for emergencies requiring immediate outlays.

3 Experimental Design

Our experiment involves a web-based financial education intervention narrowly focused on the concept of compound interest. We chose this topic for a number of reasons. First, it is associated with a well-documented behavioral bias that an intervention, if effective, would counteract. Second, it is a fundamental concept in financial decision making and most financial education courses cover it. Third, its narrowness, and the corresponding brevity of treatments in standard investment guides and employer-sponsored financial education programs, make it suitable for an intervention of limited duration.

The experiment consisted of three stages. First, subjects watched one of four educational videos, selected at random. Second, they completed incentivized valuation tasks. Finally, they took a test on compound interest, and answered survey questions concerning the decision strategies they deployed in the second stage. Performance on the test was incentivized, and subjects knew this prior to watching the educational video. Additional explanation of each stage follows; for further details, see Online Appendix A.

Education intervention. We used a video based on the section on compound interest from a popular investment guide, *The Elements of Investing: Easy Lessons for Every Investor*, by Malkiel and Ellis (2013). We selected this book because it is extremely well-exposited, widely read, and targets young adults who are beginning to think about long-term financial objectives, a group to which most of our subjects belong.

The text begins with a simple explanation of compound interest illustrated through an iterative calculation.[18] The remainder of the text consists of two components:

(i) An explanation of a simple, memorable, and potentially valuable heuristic, the rule of 72, along with five illustrative applications.[19] The rule of 72 is a method for approximating an investment's doubling period; one can also use it to approximate the growth in an investment's value over a fixed holding period. It states that the percentage interest rate on an investment multiplied by the number of periods required for its value to double equals 72 (approximately).

[18] The example is: "Stocks have rewarded investors with an average return close to 10 percent a year over the past 100 years. Of course, returns do vary from year to year, sometimes by a lot, but to illustrate the concept, suppose they return exactly 10 percent each year. If you started with a $100 investment, your account would be worth $110 at the end of the first year—the original $100 plus the $10 that you earned. By leaving the $10 earned in the first year reinvested, you start year two with $110 and earn $11, leaving your stake at the end of the second year at $121. In year three you earn $12.10 and your account is now worth $133.10. Carrying the example out, at the end of 10 years you would have almost $260—$60 more than if you had earned only $10 per year in 'simple' interest."

[19] We used this particular investment guide in part because it teaches a useful quantitative heuristic. Some investment guides and educational interventions cover this topic without offering useful quantitative tools.

(ii) Motivational material (rhetoric and exhortations). The section opens with the observation that "Albert Einstein is said to have described compound interest as the most powerful force in the universe." It provides various anecdotes concerning small investments that grew to impressive sums (in some cases millions of dollars) over long time periods. These anecdotes do not include any computations, and hence are not helpful for understanding the mechanics of compound interest. It also explicitly exhorts readers to behave frugally, asserting that "the power of compounding is why everyone agrees that saving early in life and investing is good for you," and characterizing compounding as a "miracle."

We employ a 2×2 *between subjects* design to isolate the features of the educational intervention that drive changes in test-scores, self-reported decision strategies, choices, and welfare. In our *Full* treatment, subjects viewed a video covering all of the material, both substantive and rhetorical. In our *Substance-Only* treatment, they viewed a shorter video covering all of the substantive material, but omitting exhortations and atmospheric quotes.[20] In contrast, for the *Rhetoric-Only* treatment, subjects viewed a video containing all of the rhetorical material and exhortations, as well as the introductory explanation of compound interest, but omitting all material on the rule of 72. Finally, subjects in the *Control* treatment viewed a stylistically similar video based on a section about index funds from the same investment guide. This section does not mention compound interest or the time value of money, and consequently we would not expect it to affect the types of choices that subjects were subsequently asked to make.

Subjects viewed videos of narrated slide presentations.[21] The narration was verbatim from the text (with a few minor adjustments), while the slides summarized key points. In format, the videos resemble those offered through the educational internet platform *www.khanacademy.org.* Since our study is internet-based, we took several precautionary measures to ensure that subjects were able to view the video and that they would pay attention to it. These are detailed in the Online Appendix A.

Valuation tasks Subjects performed 10 paired valuation tasks. Each task elicited an equivalent current dollar value for a reward r to be received in either 36 or 72 days. With *simple framing*, the reward was described as follows: "We will pay you \$$r$ in t days." With *complex framing*, the same reward was described in terms of a return on an initial investment, as follows: "We will invest \$$a$ at an interest rate of R% per day. Interest is compounded daily. We will pay you the proceeds in t

[20] In cases where it was impossible to remove sentences containing rhetorical material, we substituted neutral language. For instance, the first example of compounding presented in the original text is preceded by the transitional question, "Why is compounding so powerful?" In the Substance-Only-treatment, we substituted the question, "How does compounding work?"

[21] We chose this approach because existing research indicates that financial education videos are generally more effective than written text (Lusardi et al., forthcoming).

days." Subjects made two sets of choices pertaining to each future reward, one with simple framing, the other with complex framing.[22] For each frame f (which includes the description of a and R for complex framing), we elicited a subject j's immediate dollar equivalent of a payment r received in t days, $V^f_{j,r,t}$, using the iterated multiple price list method with a resolution of \$0.20 (Andersen et al., 2006).[23] We randomized the order of the valuation tasks at the subject level. Subjects were not told that some of the tasks were substantively equivalent, and they typically did *not* perform equivalent simply and complexly framed tasks consecutively.

Table 1 lists the parameters t, r, a, and R used for the paired valuation tasks. We chose time horizons of 36 and 72 days to simplify applications of the rule of 72.[24] Because our design is thereby skewed towards settings in which the substantive content of the intervention is potentially most useful, our study is biased in favor of finding *beneficial* behavioral effects. We chose values for the remaining parameters to create variation in the number of times the initial investment doubles over the investment horizon. This allows us to investigate the cause of differences between valuations for complexly and simply framed rewards: subjects who erroneously compute simple rather than compound interest make larger mistakes when the investment horizon is a larger multiple of the doubling period.

Subjects completed the paired valuation tasks at their own pace (subject to the restriction that they could not take more than 3 hours), and we recorded their response times. We intentionally placed no restriction on the use of other resources, such as calculators, the internet, or personal advice when making decisions, as subjects always have those options when making real-world decisions.[25] As detailed below, only a quarter of our subjects report using such resources when completing the incentivized test, a fraction that does not vary meaningfully across treatments. That pattern mirrors findings concerning real financial decisions (Lusardi and Mitchell, 2011).

Knowledge test and self-reports. We also gathered data to evaluate the educational intervention according to conventional metrics. Many studies have used tests of knowledge and understanding (e.g. Jump$tart Coalition for Personal Financial Literacy, 2006; Mandell, 2009; Mandell and Klein, 2009; Carpena et al., 2011; Heinberg et al., 2014; Lusardi et al., forthcoming; Walstad, Rebeck and Macdonald, 2010; Council for Economic Education, 2006; Collins, 2010). Accordingly, we administered

[22] We chose the parameters of the tasks so that the complexly framed version yielded the same future payment as the simply framed version according to the rule of 72. Since that rule is an approximation, future values actually differ by small amounts between the two frames.

[23] Throughout, we set $V^f_{j,r,t}$ equal to the midpoint of the pertinent interval. For further details, see Online Appendix A.

[24] We used two different time frames so subjects would face a greater variety of decision problems, and hence would be less likely to consider successive problems highly similar.

[25] This feature differentiates our study from most of the literature on the effects of financial education (Hastings, Madrian and Skimmyhorn, 2013). An exception is Levy and Tasoff (2016) who also conduct an internet-based study.

Future Reward r	Investment Amount a	Daily Interest Rate R	Number of Doublings
	Duration: 72 days		
\$20	\$10	0.01	1
\$18	\$4.5	0.02	2
\$16	\$2	0.03	3
\$14	\$0.9	0.04	4
\$12	\$2	0.025	2.5
	Duration: 36 days		
\$20	\$10	0.02	1
\$18	\$4.5	0.04	2
\$16	\$2	0.06	3
\$14	\$0.9	0.08	4
\$12	\$2	0.05	2.5

Table 1: Decision problems. *Number of doublings* is the number of times the initial investment doubles over the investment horizon according to the rule of 72. Final amounts are calculated using the rule of 72. Exact final amounts differ by no more than \$0.80, except for the 4% interest rate over 72 days, where the rule understates the future value by \$1.16. Our analysis controls for these differences.

an incentivized test consisting of the five questions about compound interest listed in Table 2, as well as five questions about the material covered in the video shown to the control group.[26]

Previous studies have also examined self-reported decision strategies (for instance Heinberg et al., 2014; Lührmann, Serra-Garcia and Winter, 2012; Carlin, Jiang and Spiller, 2014). In the final stage of the experiment, we asked subjects whether they had used the rule of 72 in the complexly framed problems, and whether they had used it in the simply framed problems. We also elicited the number of complexly framed valuation tasks for which subjects explicitly calculated the future value of the investment, and asked whether they obtained help when taking the test on compound interest.[27]

4 Implementation and Preliminary Analysis

We conducted our experiment through the online labor market Amazon Mechanical Turk (AMT).[28] An important feature of this population is that the typical member has a poor understanding of compound interest. Also, this group resembles the target populations for many financial education

[26]The test questions for the material in the control video are available upon request. We randomized the order of all ten test questions at the subject level. Subjects knew that their test results and choices in the paired valuation tasks would determine their rewards with 25% and 75% probabilities, respectively. For the test results, they received \$1 for each question they answered correctly.

[27]The questionnaire also addressed a small number of additional issues.

[28]An advantage of conducting the experiment online is that it mirrors many real-world financial decisions, which have steadily migrated to internet platforms.

Q1. If the interest rate is 10% per year (interest is compounded yearly), how many years does it take until an investment doubles?
7 years, 7.2 years, 7.4 years, 7.8 years, 8 years

Q2. If somebody tells you an investment should double in four years, what rate of return (per year) is he promising?
15%, 16%, 17%, 18%, 19%, 20%

Q3. If the interest rate is 7% per year (interest is compounded yearly), about how long does it take until an investment has grown by a factor of four (i.e. is four times as large as it was originally)?
About 5 years to *about 40 years*, in steps of 5 years.

Q4. Paul had invested his money into an account which paid 9% interest per year (interest is compounded yearly). After 8 years, he had $500. How big was the investment that Paul had made 8 years ago?
$200 to *$400* in steps of $10

Q5. If an investment grows at 8 percent per year (interest is compounded yearly), by how much has it grown after 4 years?
By 30%, to *by 40%* in steps of one percentage point.

Table 2: Test questions. Questions were presented in random order and intermingled with the questions concerning material covered in the Control video.

programs in terms of demographic characteristics such as age and income. Broadly, experience to date indicates that AMT provides a useful and reliable platform for many types of behavioral research in the social sciences (Horton, Rand and Zeckhauser, 2011; Mason and Suri, 2012; Peysakhovich, Nowak and Rand, 2014).

We ran eight sessions with a total of 504 subjects during April and May 2014, all on weekday mornings. We restricted participation to subjects who reside in the US and are at least 18 years of age. Subjects logged into our study from the AMT worker interface. They were welcomed by a two-and-a-half minute video recording of one of the authors (Bernheim), who vouched that we would pay subjects exactly the amount we promised them within at most two days of the promised date.[29] Before participating in the main stages of the experiment, subjects completed an unincentivized questionnaire concerning demographics, as well as a standard battery of five questions designed to assess financial literacy.[30]

[29]The video invited subjects to click a link to the author's homepage so they could verify the authenticity of the video. It also provided a link to the homepage of a graduate-student co-author (Ambuehl) in case they felt uncomfortable contacting and inconveniencing a professor.

[30]This test of financial literacy originated with Lusardi and Mitchell (2009) and van Rooij, Lusardi and Alessie (2011), and has been used in many other studies (Lusardi and Mitchell, 2014). We reproduce the five questions in the Online Appendix Table A.1. It is standard practice to administer this test without incentivization.

The average length of a session was 62 minutes (s.d. 22 minutes). Attrition was negligible and unrelated to the treatments.[31] On average, subjects earned \$22.86, including a fixed \$10 participation fee; earnings ranged from a low of \$10 to a high of \$30.47. In comparison, AMT participants typically earn about \$5 per hour (Mason and Suri, 2012).

Multiple switching. Any subject with coherent preferences will switch her choice from the immediate payment to the future reward at most once within a single price list. We did *not* impose this restriction on our subjects, but instead informed them that "most people begin a decision list by preferring the option on the left and then switch to the option on the right." As a result, 7.7% of subjects (39 of 504) switched two or more times in at least one price list, and this number does not significantly differ across treatments ($p = 0.85$). In laboratory studies of risky choices by undergraduate subjects (such as Holt and Laury, 2002), the comparable figure typically falls in the range of 10 to 15%. Following the usual convention (see, for example, Harrison et al., 2005), we focus attention on the 455 subjects who respected monotonicity .

Demographics. While our subjects are not highly representative of the US population, neither are they highly unusual. On average, our sample is somewhat poorer, better educated, and more likely to live in larger households than the average US citizen. While our sample mirrors the general population with respect to the prevalence of full-time employment, the fraction of respondents who describe themselves as working part-time is twice as high. Perhaps because we recruited our subjects through the internet, our sample over-represents males, young adults, whites, urban residents, and people who have never been married. The level of financial literacy slightly exceeds that found in other studies of US subjects (see Lusardi and Mitchell, 2009, and Lusardi, 2011). Online Appendix B.1 provides additional details.

Randomization into treatments was successful. Of the 34 F-tests we performed to assess the differences in demographic characteristics across treatments (one for each characteristic), two are significant at the 5%-level, and two more are significant at the 10% level. These figures are well within the expected range. Online Appendix B.1 also includes these tests.

Attention. A concern with studies conducted on internet platforms is that some subjects may pay insufficient attention to the experimental tasks. We motivated subjects to attend by providing monetary incentives that were large relative to the wages for which they ordinarily work, and by emphasizing

[31] Only four subjects who reached the stage at which they may have viewed a treatment video failed to complete the study. A larger number of subjects quit before reaching that stage, but that type of attrition is necessarily independent of the treatment, and hence largely innocuous; also, there is no reason to think that the pre-attrition sample is more representative of the general population than the post-attrition sample. Technical glitches may be responsible for both kinds of attrition. For example, a small number of subjects contacted us to report that the video failed to load on their computers.

the broader value of understanding the material covered in the videos. Several findings suggest that we were successful. First, choice patterns are coherent, both with respect to time preferences, and with respect to our educational interventions. Second, the extremely low rate of attrition (mentioned above) indicates that subjects were highly engaged. Indeed, many subjects provided us with unsolicited positive feedback concerning the educational interventions. Third, we obtain similar results when subjects who exhibited either unusually noisy or unresponsive behavior – the likely hallmarks of inattention – are dropped from the sample; see Online Appendix C.2. Finally, we take a small degree of reassurance from the fact that, when completing the exit survey, the overwhelming majority of subjects reported paying the highest level of attention to the video and to their choices.

Baseline discounting and exponential growth bias. The extent to which subject j discounts a reward r in a decision task with time horizon t and frame $f \in \{simple, complex\}$ is given by

$$\delta^f_{j,r,t} = \frac{V^f_{j,r,t}}{r} \tag{1}$$

We also refer to this quantity as the subject's *normalized valuation.* Focusing on the Control condition, the average normalized valuations with simple framing (that is, discount factors) are 0.767 and 0.706 for tasks with 36 and 72 day horizons, respectively.[32] There is also significant exponential growth bias: normalized valuations with complex framing are lower than with simple framing by an average of 13.3 percent of the promised reward.

5 Conventional Outcome Measures

As noted in Sections 1 and 3, studies that evaluate financial education interventions frequently focus on financial literacy, self-reported decision strategies, and directional changes in behavior. In this section, we show that our intervention appears to be successful according to these conventional outcome measures.[33] In section 6 we then show that these measures fail to detect crucial deficiencies.

Column 1 of table 3 shows the effects of our treatments on subjects' test scores for the five questions pertaining to compound interest. In the Control condition, the average subject answers just under two of five, or 39%, of the questions correctly. The Full intervention increases the average score dramatically, by roughly 1.4 additional correct answers, or equivalently by 29 percentage points, to

[32] Thus our typical subject discounts future payments rather heavily. A longer horizon results in greater discounting, but the relative magnitudes of these rates across horizons are inconsistent with exponential discounting. These patterns are common in studies that elicit time preferences over short horizons (Frederick, Loewenstein and O'Donoghue, 2002). They do not, however, reflect conventional present-bias, because our subjects expect to receive all payments with a one-to-two-day lag.

[33] All results reported in this section are robust with respect to various statistical controls and alternative specifications. For details, see Online Appendix C.

68%. When the rhetoric is removed from the intervention (the Substance-Only treatment), the effect is only slightly smaller, and the difference is not statistically significant. In contrast, when material on the rule of 72 is removed (the Rhetoric-Only treatment), the average score improves by only 0.5 additional correct answers, or equivalently 10 percentage points.[34] Thus, according to standard measures, the interventions that include substantive material are highly effective at promoting financial literacy.[35]

VARIABLES	(1) Test score compounding	(2) Test score control	(3) External help	(4) Uses rule in complex framing	(5) Uses rule in simple framing	(6) Explicit calculation	(7) $100 \times \delta^{c}_{j,r,c}$
Level in Control	1.963***	3.284***	0.220***	0.128***	0.092**	6.404***	58.95***
	(0.139)	(0.103)	(0.042)	(0.040)	(0.039)	(0.354)	(2.272)
Treatment effects							
Full	1.442***	-1.058***	-0.013	0.579***	0.172***	1.738***	14.31***
	(0.197)	(0.146)	(0.059)	(0.056)	(0.055)	(0.504)	(3.427)
Substance-Only	1.271***	-1.339***	0.061	0.637***	0.260***	1.737***	4.021
	(0.189)	(0.140)	(0.057)	(0.054)	(0.053)	(0.482)	(3.285)
Rhetoric-Only	0.492**	-1.079***	0.066	0.104*	0.060	0.418	18.59***
	(0.195)	(0.144)	(0.058)	(0.056)	(0.054)	(0.497)	(3.595)
$P(\beta_{Substance}{=}\beta_{Rhetoric})$	0.000	0.062	0.937	0.000	0.000	0.006	0.000
$P(\beta_{Full}{=}\beta_{Rhetoric})$	0.000	0.885	0.184	0.000	0.040	0.009	0.259
$P(\beta_{Substance}{=}\beta_{Full})$	0.368	0.047	0.196	0.285	0.099	0.999	0.003
P(joint insignificance)	0.000	0.000	0.400	0.000	0.000	0.000	0.000
Observations	455	455	455	455	455	455	4,550
Number of subjects	455	455	455	455	455	455	455

Table 3: Conventional outcome measures. The dependent variable in columns 1 - 7 are, respectively, the mean number of test questions answered correctly (1 to 5), the self-reported answer to whether the subjects used external help in the test, the answer to the question whether the rule of 72 was used in the complexly framed problems, the answer to the question whether the rule of 72 was used in the simply framed problems, the self-reported number of complexly framed problems (out of 10) for which the subject explicitly calculated the future reward, and the normalized valuation for complexly framed decision tasks. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

These improvements in performance on test questions pertaining to compound interest are not due to effects of the Full, Substance-Only, and Rhetoric-Only videos on general motivation. If they were, we would find comparable effects for subjects' scores on the five test questions pertaining to topics covered in the Control video. On the contrary, as shown in Column 2, the Control video increases the average score on this portion of the test by more than one additional correct answer (over 20%)

[34] The fact that there is still a gain is not surprising given that Rhetoric-Only treatment, unlike the Control treatment, includes a simple explanation of compound interest, illustrated through an iterative calculation.

[35] See Online Appendix C.1 for the effects on individual test questions.

relative to all three treatments. We conclude that subjects learn the substantive material contained in whichever video they view.

A natural concern is that education may simply displace the use of reference materials or reliance on knowledgable friends. Such displacement could dampen the effects of the interventions on test scores and choices. Column 3 shows that the various educational interventions do not affect the (self-reported) extent to which subjects employ external help.

Subjects report operationalizing the knowledge they acquire from the substantive interventions, as Column 4 shows. Only 13% of subjects in the Control report using the rule of 72 when making complexly framed choices. In sharp contrast, the corresponding figure exceeds 70% for the Full and Substance-Only treatments. Somewhat surprisingly, we also find an increase – albeit a much smaller one – for the Rhetoric-Only treatment.[36] As shown in Column 5, we find a qualitatively similar pattern for self-reported operationalization of the rule of 72 in simply framed choices; however, the frequencies and treatment effects are all considerably smaller than for complexly framed choices. Because subjects may report using the rule of 72 in simply framed problems for a variety of reasons, this finding is not entirely unexpected.[37]

In principle, the increased use of the rule of 72 could crowd out other types of calculations, such as iterative computations, applications of the compound interest formula, or (inappropriate) evaluations of simple interest. Depending on the nature of the displaced approach, such crowding out could dampen the effect of education on test scores and behavior. In fact, Column 6 shows that the Full and Substance-Only interventions significantly increase the average number of complexly framed decision tasks for which subjects report making explicit calculations, from roughly 6.4 to 8.1 out of 10 (i.e., by approximately 27%). For the Rhetoric-Only treatment, the effect is much smaller and statistically insignificant. Thus, the educational interventions do not simply increase (self-reported) reliance on the rule of 72 by migrating subjects from other methods of explicit calculation.

Next we turn to the effects of financial education on behavior. Many studies draw informal inferences concerning the success of these types of interventions by asking whether they directionally counteract presumed biases. For instance, financial education interventions are often deemed successful if they increase contributions to retirement savings accounts. For the types of decisions we examine in this study, it is well-established that people on average underestimate the power of compound interest, a phenomenon known as *exponential growth bias* (see the references cited in Section 1). Consequently, following the approach adopted in the literature, one would deem an intervention

[36]There are two possible explanations for this finding. One is that some subjects already know the rule of 72 but apply it only when they are sufficiently motivated. The other is that rhetorical exhortation motivates subjects to misrepresent their knowledge and use of the rule.

[37]Subjects may apply the rule inappropriately, they may discount future rewards to the present at a market interest rate, or they may misrepresent their actual decision processes.

potentially welfare-improving if it leads subjects to value investments involving compound interest (our complexly framed rewards) more highly.

Column 7 shows the effects of the various treatment videos on normalized valuations for complexly framed tasks. According to the table, the Full video increases valuations for complexly framed choices by a 14.31 percentage points relative to the Control video, and the effect is highly significant. Furthermore, given the magnitude of the exponential growth bias documented in the existing literature, the size of the average treatment effect raises no concerns about systematic overcorrection.[38]

Taken at face value, the preceding results suggest that the Full intervention has the right effects for the right reasons. It successfully increases performance on an incentivized knowledge test and, as one would hope, this increase results from the substantive elements of the intervention rather than from motivational rhetoric. Moreover, subjects report operationalizing their newly obtained knowledge in their decisions, and there is no indication that the use of new quantitative tools crowds out reliance on other resources or computational methods. Finally, valuations in complexly framed tasks change in a direction that counteracts a known bias (which we have verified for this sample), and the change does not appear to be excessive on average. Based on these results, one would expect to find that the Full intervention unambiguously improves the quality of financial decision making, and that this effect is driven by substantive material rather than rhetoric.

As we will see, the results presented in the next section paint a much different picture, which demonstrates the value of formally assessing the quality of financial decision making, as we do in our work. A closer examination of the regression in the final column of Table 3 alerts us to the source of the problem: the estimated effect on valuations in complexly framed tasks for the Substance-Only treatment (4.02 percentage points) is statistically indistinguishable from zero and significantly smaller than that of the Full treatment (14.31 percentage points, $p = 0.003$). In contrast, the estimated effect for the Rhetoric-Only treatment (18.59 percentage points) is actually larger than that of the Full treatment, and we do not reject equality ($p = 0.259$). Accordingly, despite demonstrable effects of substantive instruction on comprehension as well as subjects' statements concerning their proclivities to operationalize substantive knowledge in their decisions, *the behavioral effects of the Full treatment are traceable almost entirely to motivational rhetoric rather than substance.*

[38]Stango and Zinman (2009) posit that subjects assess future value (FV) based on the magnitude of an initial investment (I) and the interest rate (i) according the formula $FV = I \times (1+i)^{\theta t}$. They estimate this equation for each member of their subject pool. The median estimate of θ is 0.8 (see their footnote 24). Given the tasks in our experiment, a subject with $\theta = 0.8$ underestimates future values on average by a factor of 0.71. Assuming that current valuation varies proportionately with the magnitude of the future receipt, the elimination of exponential growth bias would increase the average current valuation by 40.1% (because $1/0.71 = 1.401$). In contrast, the Full treatment increases the mean valuations for complexly framed tasks by $14.31/58.95 = 24.3\%$. Thus it appears from this calculation that the Full treatment did not cause subjects to overcorrect on average.

6 The Quality of Decision Making

Next we assess the effects of our educational interventions on the quality of decision making using the approach developed in Section 2. We demonstrate that, despite the generally encouraging results of the previous section, the Full intervention fails to make subjects better off on average. Additional results link this finding to the role of motivational rhetoric.

VARIABLES	(1) $100 \times \delta^s_{j,r,c}$	(2) $100 \times d_{j,r,c}$	(3) $100 \times C_e$	(4) $100 \times C_m$
Level in Control	72.26***	-13.31***	11.69***	24.45***
	(2.089)	(2.221)	(1.232)	(1.633)
Treatment effects				
Full	0.402	13.91***	0.155	-1.584
	(2.99)	(3.332)	(2.035)	(2.386)
Substance-Only	0.018	4.002	-1.461	-2.436
	(2.913)	(2.961)	(1.669)	(2.172)
Rhetoric-Only	5.368*	13.22***	-2.546	-4.651**
	(2.975)	(2.952)	(1.700)	(2.155)
$P(\beta_{Substance} = \beta_{Rhetoric})$	0.069	0.001	0.505	0.270
$P(\beta_{Full} = \beta_{Rhetoric})$	0.010	0.827	0.177	0.171
$P(\beta_{Substance} = \beta_{Full})$	0.897	0.002	0.413	0.706
P(joint insignificance)	0.202	0.000	0.390	0.178
Observations	4,550	4,550	4,550	4,550
Number of subjects	455	455	455	455

Table 4: Results pertaining to the quality of decision making. $\delta^f_{j,r,t}$ is subject j's normalized valuation for reward r to be received at time t when presented in frame f. $d_{j,r,c} = \delta^c_{j,r,t} - \delta^s_{j,r,t}$ is the framing distortion. If subject j underestimates compound interest, $d_{j,r,c} < 0$. Subject j's expected and maximal welfare losses from characterization failure are proportional to $C_e = (d_{j,r,c})^2$ and $C_m = |d_{j,r,c}|$, respectively. Standard errors clustered by subject. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Simply framed valuations. We start by verifying that subjects' valuations for simply framed opportunities are largely invariant with respect to the educational interventions. As explained in section 2, the welfare interpretation of our financial competence measures presupposes this stability property. Column 1 of Table 4 shows that, for normalized valuations in simply framed tasks, the estimated effects of the Full and Substance-Only interventions are close to zero and statistically insignificant. While the corresponding effect of the Rhetoric-Only condition is somewhat larger, it is less than one-third the size of its counterpart in the regression for complexly framed valuations (column 7 of Table 3), and it lacks statistical significance at the 5% level.

Framing distortions. Next we investigate the effect of our educational interventions on the degree of exponential growth bias. To this end, we define the *framing distortion* as the difference between the normalized complex and simple valuations for the same task: $d_{j,r,c} = \delta^c_{j,r,t} - \delta^s_{j,r,t}$. (Note that we do *not* take the absolute value.) An individual who underestimates (overestimates) the power of compound interest will exhibit $d_{j,r,t} < 0$ ($d_{j,r,t} > 0$). As we mentioned in Section 4, our subjects exhibit substantial exponential growth bias in the Control condition. Indeed, column 2 of table 4 shows subjects' normalized valuations are lower with complex framing than with simple framing by, on average, 13.31 percentage points (measured relative to the promised reward).

In light of the fact that our Full intervention increases valuations for complexly framed tasks but not for simply framed tasks, one should not be surprised to learn that it reduces the magnitude of the average framing distortion. Even so, the extent of the reduction is striking. According to column 2, the average value of $d_{j,r,c}$ falls by 13.91 percentage points, leaving a gap of only 0.6 percentage points (s.e. $= 2.48$), thereby effectively eliminating exponential growth bias on average. While this result is in line with the encouraging findings of the previous section, we emphasize that the effect flows almost entirely from motivational rhetoric rather than the substantive elements of instruction. In particular, the estimated effect on the mean framing distortion for the Substance-Only treatment (4.00 percentage points) is statistically indistinguishable from zero, and significantly smaller than that of the Full treatment (13.91 percentage points, $p = 0.002$). In contrast, the estimated effect for the Rhetoric-Only treatment (13.22 percentage points) is almost identical to that of the Full treatment, and we do not reject equality ($p = 0.827$).

Because the elimination of the average framing distortion results from motivational rhetoric rather than substance, one suspects that these averages may mask many inappropriate subject-level responses. To investigate this possibility, we examine the cumulative distribution of $d_{j,r,t}$ for each treatment; see Figure 1. While subjects usually exhibit exponential growth bias in the Control treatment ($d_{j,r,t} < 0$ in roughly 65% of tasks), they also *overestimate* compound interest with reasonably high frequency ($d_{j,r,t} > 0$ in roughly 35% of tasks). Moreover, much of the observed variation in the framing distortion reflects individual-level heterogeneity rather than task-specific noise.[39]

[39] The Cronbach-α statistics for $d_{j,r,t}$ show that subjects who underestimate (overestimate) compound interest in some decisions tend do so in all decisions, and by comparable amounts. The values of the statistic are 0.92, 0.92, 0.94, and 0.95 for the Control, Full, Substance-Only and Rhetoric-Only treatments, respectively. These values compare favorably with the standard benchmark of 0.8, indicating a high level of individual consistency.

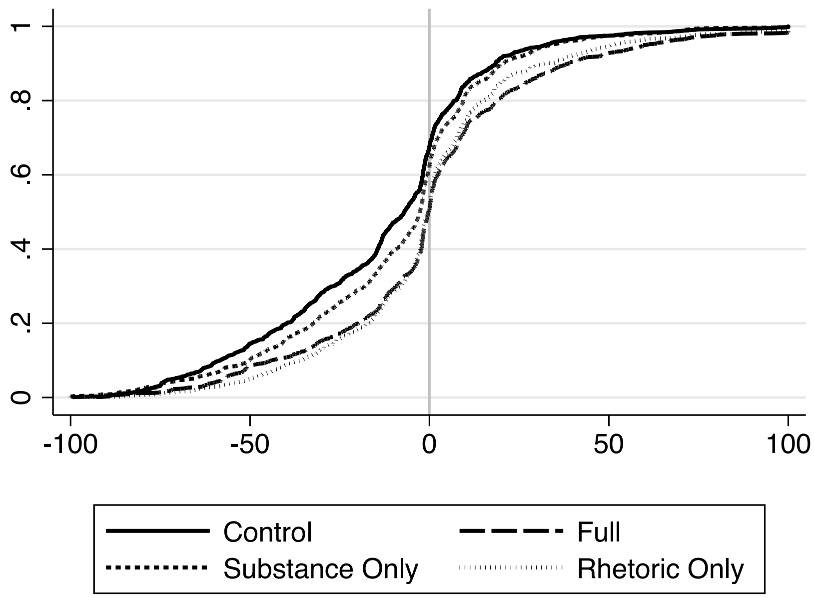

Figure 1: C.D.F. of framing distortion, by treatment. For better visibility, the graph is truncated at -100 and at 100.

An effective intervention would bring valuations for equivalent complexly and simply framed problems more closely in line. For subjects who underestimate compound interest, it would increase valuations complexly framed tasks. For subjects who overestimate compound interest, it would *decrease* these valuations. As a result, the two CDFs would *cross* at $d = 0$, and the distribution for the intervention would be more tightly concentrated around zero.

Instead, *the Full intervention shifts the entire CDF to the right.* In other words, it generally increases valuations for complexly framed tasks irrespective of whether the subject initially underestimates or overestimates compound interest. This indiscriminate effect helps in some instances but hurts in others.[40] The Rhetoric-Only treatment yields a similar shift in the CDF, while the Substance-Only treatment has a much smaller effect. These findings are consistent with the hypothesis that behavioral effects of the Full intervention primarily reflect motivational elements of instruction rather than substantive elements, and that consequently they bear no systematic relation to the appropriate response.

Financial competence and welfare. In light of the preceding findings, the effects of the Full intervention on welfare are unclear. On the one hand, it significantly enhances financial literacy,

[40]We are not alone in finding that some people overestimate compound interest; see, for example, Goda et al. (2015) and Levy and Tasoff (2016).

induces people to operationalize their knowledge in their decisions without reducing reliance on other resources (according to self-reports), increases the frequency with which people report using decision strategies that involve explicit calculations, and brings average complexly framed valuations into almost perfect alignment with average simply framed valuations. On the other hand, its behavioral effects are driven almost entirely by its motivational elements rather than its substantive elements, and as a result its impact is largely indiscriminate (that is, unrelated to the initial framing distortion).

To determine whether an intervention improves or reduces welfare, we examine its effects on financial competence, measured as either $C_e = (\delta^s_{j,r,t} - \delta^c_{j,r,t})^2$ or $C_m = |\delta^s_{j,r,t} - \delta^c_{j,r,t}|$. These measures are always non-negative, and higher values imply lower competence. Columns 1 and 2 of Table 4 show how our interventions affect them. There is no evidence that the Full treatment benefits subjects by improving their financial competence on average. The point estimates for the effects of the Substance-Only and Rhetoric-Only treatments on C_e and C_m are a bit larger in magnitude, but only one of the four is statistically significant (the Rhetoric-Only treatment effect on C_m).

It is natural to wonder whether our findings concerning financial competence are attributable to a mismatch between the difficulty of the valuation problems and the depth of the material covered in the Full intervention. To investigate that possibility, we reexamine the effects of the various interventions on welfare, differentiating between tasks according to the difficulty of applying the rule of 72. The rule is easiest to apply when the investment in question doubles only once over the time horizon, more difficult to apply when it doubles an integer number of times, and most difficult to apply when it doubles a non-integer number of times. Accordingly, we re-estimate the basic specification from Table 4 separately for valuation tasks with a single doubling, two to four doublings, and 2.5 doublings. Results appear in columns 3 - 5, respectively, of table 5.

If the ease of applying the rule of 72 improves the success of interventions that teach it, we should see systematic differences in the *relative* welfare effects of the substantive and Rhetoric-Only interventions across these three categories of valuation tasks.[41] Thus, in table 5, we would expect to find that the *difference* between the effect of the Full (or Substance-Only) treatment and the Rhetoric-Only treatment decreases as we move from column 3 to columns 4 and 5, thereby increasing the difficulty of applying the rule. In fact, no such pattern is observed. We cannot reject the hypothesis that the difference between the welfare effects of the Full and Rhetoric-Only treatments is the same for all three classes of valuation tasks ($p > 0.10$ for all pairwise comparisons). The same is true of the difference between the welfare effects of the Substance-Only and Rhetoric-Only treatments ($p > 0.10$

[41] Notice that our focus here is on the relationship between *relative* welfare effects and the difficulty of applying the rule of 72. For any given treatment, the *absolute* welfare effects may vary with that degree of difficulty for other reasons. For example, difficulty is associated with the number of doublings, which in turn is associated with initial degree of exponential growth bias. Mechanically, any fixed increase in valuation is more likely to be welfare enhancing when the initial bias is greater.

for all pairwise comparisons).[42] Thus, one cannot attribute the poor performance of our substantive interventions in terms of welfare to the difficulty of applying the rule of 72 in our valuation tasks.

VARIABLES	(1) $100 \times C_e$	(2) $100 \times C_e$	(3) $100 \times C_e$	(4) $\tau^c_{j,r,t}$	(5) $\tau^s_{j,r,t}$
Doublings	1	[2 3 4]	2.5		
Level in Control	-7.036***	-13.00***	-12.42***	50.81***	22.41***
	(1.435)	(1.409)	(1.909)	(2.675)	(1.073)
Treatment effects					
Full	-0.302	-1.578	5.808	2.904	-0.768
	(1.889)	(2.079)	(3.983)	(4.471)	(1.724)
Substance-Only	-1.663	-1.914	0.102	19.51***	-0.428
	(1.616)	(1.919)	(2.688)	(7.069)	(1.472)
Rhetoric-Only	-2.428	-4.601**	3.502	10.08**	0.941
	(1.589)	(1.829)	(3.310)	(4.599)	(2.197)
$\beta_{Full} - \beta_{Rhetoric}$	1.946	3.023	2.306		
$\beta_{Substance} - \beta_{Rhetoric}$	0.765	2.687	-3.400		
Observations	910	2,730	910	4,550	4,550
Subjects	455	455	455	455	455

Table 5: Problem difficulty and response times. Columns 1 - 3 show the effect on average welfare for complexly framed decision tasks that differ according to the number of times the investment doubles over its life. Columns 4 and 5 show the effect of the treatments on mean response times for the complexly and simply framed problems, respectively. Standard errors clustered by subject. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Decision times. An examination of decision times corroborates some the assumptions made in Section 2 as well as various inferences we have drawn from our analysis of valuations. We derive this corroboration from the regressions in the final two columns of Table 5, which pertain to decision times in complexly and simply framed valuation tasks, $\tau^c_{j,r,t}$ and $\tau^s_{j,r,t}$, respectively. Several notable conclusions follow from these regressions.

First, on average, valuation tasks with complex framing take subjects nearly three times as long to complete than those with simple framing (59 seconds versus 22 seconds, $p < 0.001$). This finding is consistent with our premise that simply framed tasks are transparent and easily evaluated, while complexly framed tasks require additional cognitive effort, likely because subjects try to "translate" from complex to simple framing.

Second, decision times are sensitive to the educational interventions for complexly framed tasks, but not for simply framed tasks. This pattern is expected given the sensitivity of complexly framed

[42] We note that ($\beta_{Substance} - \beta_{Rhetoric}$) is significantly different across non-integer and integer doublings ($p < 0.05$). However, the actual sign of this difference is opposite the hypothesized sign.

valuations and the insensitivity of simply framed valuations to the same interventions, as well as the relative frequencies with which subjects report using information from the videos when performing the two types of tasks. Together, these findings provide a solid foundation for our assumption that the interventions change the way subjects think about and assess opportunities that are complexly framed, but not ones that are simply framed. As we hypothesized, education appears to alter the "translation" from complex to simple framing.

Third, the pattern of effects for complexly framed tasks corroborates our inferences about the role of motivational rhetoric. The Substance-Only intervention has the largest effect, increasing average decision times by 19.5 seconds relative to the Control, or roughly 40%. Thus, this intervention appears to alter the way subjects think about complexly framed opportunities, even though it does not produce much of a systematic shift in *average* valuations. (Below, we show that it *does* change the manner in which subjects value these opportunities, but the effect is directionally haphazard.) In contrast, the impact of the Full treatment is small and statistically insignificant. Thus, the provision of substantive information appears to induce greater effort and deliberation, but the addition of simplistic rhetorical assertions concerning the power of compound interest seem to negate that effect, perhaps because they point to a less cognitively demanding heuristic.

Reliance on simple interest calculations. According to previous research, many people project investment values based on linear rather than exponential growth – in other words, according to simple interest (Eisenstein and Hoch, 2007; McKenzie and Liersch, 2011). As we show next, *all* of our interventions – including the Substance-Only video, for which the effect on average valuations is minimal – render this misconception less common. Thus, the problem is not one of intellectual stubbornness. Rather, the interventions apparently migrate subjects to other similarly inappropriate methods of making choices.

We estimate the frequency with which subjects employ simple interest calculations as follows. Let $FV_{r,t}^{SI}$ and $FV_{r,t}^{CI}$ denote the future value of an investment calculated according to simple and compound interest, respectively. Then $\frac{FV_{r,t}^{SI}}{FV_{r,t}^{CI}}$ represents the degree to which simple interest understates the investment's true value. If subject j's choices are guided by the simple interest formula, then this ratio should correlate with his valuation ratio, $\frac{V_{j,r,t}^{c}}{V_{j,r,t}^{s}}$. In contrast, if j's choices are consistent with correct compounding, then his valuation ratio should equal one.

Formally, we estimate the following regression model:

$$\frac{V_{j,r,t}^{c}}{V_{j,r,t}^{s}} = \sum_{\tau \in T} \left[\beta_0^{\tau} + \beta_1^{\tau} \frac{FV_{r,t}^{SI}}{FV_{r,t}^{CI}} \right] \mathbb{I}_j(\tau) + \epsilon_{j,r,t} \tag{2}$$

where $T = \{Control, Full, Substance, Rhetoric\}$ is the set of all treatments, and $\mathbb{I}_j(\tau)$ is an indicator function that equals 1 if subject i is in treatment τ.[43] In this specification, β_1^τ gauges the prevalence of simple interest calculations. Suppose for example that all subjects compute future value according to either the simple or compound interest formula. Then $\beta_0^\tau + \beta_1^\tau = 1$, and we can interpret β_1^τ as the fraction of decisions that are consistent with simple rather than compound interest calculations in treatment τ. In the extreme, if all subjects correctly calculate future value, we would find $\beta_0^\tau = 1$ and $\beta_1^\tau = 0$, and if all subjects use the simple interest formula, we would find $\beta_0^\tau = 0$ and $\beta_1^\tau = 1$.

We estimate model (2) pooling data for all of our subjects, as well as separately for subjects with high and low financial literacy, as measured by the three questions concerning the time value of money that were included in the unincentivized financial literacy test administered at the start of the experiment. In each case, we pool data across all valuation tasks.[44] Here we use median regression because the distribution of the dependent variable is highly skewed due to the presence of observations with values of $V^s_{j,r,t}$ close to zero.

Results appear in Table 6. According to our basic specification, roughly 30% of the Control group's complexly framed decisions are made using the simple interest formula. That method appears to be far more prevalent among those with low financial literacy (49%) than among those with high financial literacy (20%). The Substance-Only treatment reduces reliance on simple interest calculations to roughly 9% overall (29% and 6% for those with low and high financial literacy, respectively). Notably, both the Full and Rhetoric-Only treatments essentially eliminate dependence on simple interest calculations for both groups (though the effect of the Rhetoric-Only treatment on subjects with low financial literacy is estimated imprecisely). Hence, *all* of our treatments successfully discourage reliance on the logic of simple interest.

For all three specifications and every treatment group, $\beta_0^\tau + \beta_1^\tau$ is extremely close to unity, suggesting that our model is well-specified.[45] Absent other evidence, one might therefore be tempted to conclude that subjects make either simple interest or (correct) compound interest calculations, and that the interventions successfully push them toward the latter. However, in light of our findings concerning welfare, it is clear that, even though all of the interventions discourage the use of the simple interest

[43] Note that the dependent variable, $\frac{V^c_{j,r,t}}{V^s_{j,r,t}}$, is likely independent of subject i's time preferences: If subject i perceives future values $FV^f_{j,r,t}$ in frame f, and $V^f_{j,r,t} = \tilde{\delta} FV^f_{j,r,t}$, then $\frac{V^c_{j,r,t}}{V^s_{j,r,t}}$ is independent of $\tilde{\delta}$.

[44] In particular, our regressions employ data for valuation tasks with both 36 and 72 day horizons. As discussed elsewhere in this section, there is reason to think that subjects may be more likely to compute compound interest with 72 day horizons, at least in the treatments that teach the rule of 72. If the time horizon were systematically related to the values of $\frac{FV^{SI}_{r,t}}{FV^{CI}_{r,t}}$, our estimates of model (2) could confound the effects of the future value ratio with the effects of the time horizon. This is not a problem, however, because we have chosen the parameters of the valuation tasks so that the values of $\frac{FV^{SI}_{r,t}}{FV^{CI}_{r,t}}$ are the same for both time horizons. In any case, as shown below, the time horizon does not appear to have much of an effect on the valuation ratio in practice.

[45] We fail to reject the hypothesis that $\beta_0^\tau + \beta_1^\tau = 1$ in all cases with $p > 0.3$.

formula, they do *not* succeed in fostering the correct calculation of compound interest in the context of decisions that implicate preferences.

VARIABLE	(1)	(2) $V^c_{j,r,t}/V^s_{j,r,t}$	(3)
Sample	all	high FL	low FL
$\beta_1^{Control}$	0.304***	0.197**	0.489**
	(0.100)	(0.0908)	(0.212)
β_1^{Full}	0.00928	0.0133	-0.00397
	(0.0309)	(0.0269)	(0.134)
$\beta_1^{Substance}$	0.0883**	0.0600*	0.294*
	(0.0380)	(0.0348)	(0.178)
$\beta_1^{Rhetoric}$	0.0234	0.000435	0.0821
	(0.0329)	(0.0316)	(0.108)
$\beta_0^{Control}$	0.721***	0.814***	0.527**
	(0.0849)	(0.0797)	(0.208)
β_0^{Full}	0.993***	0.994***	0.984***
	0.0227	0.0191	0.102
$\beta_0^{Substance}$	0.906***	0.930***	0.730***
	0.0298	0.0262	0.149
$\beta_0^{Rhetoric}$	0.983***	1.000***	0.926***
	0.0199	0.0169	0.0823
$P(\beta_1^{Control} = \beta_1^{Full})$	0.00489	0.0527	0.0496
$P(\beta_1^{Control} = \beta_1^{Substance})$	0.0436	0.160	0.481
$P(\beta_1^{Control} = \beta_1^{Rhetoric})$	0.00771	0.0412	0.0880
Observations	4,550	2,920	1,630
Subjects	455	292	163

Table 6: Use of simple interest formula. High and low financial literacy (FL) are measured by the three questions concerning the time value of money that were included in the unincentivized test administered at the start of the experiment. Estimated using median regression. Standard errors clustered by subject. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Robustness Online Appendix C details a wide range of robustness analyses. First we show that our findings are not sensitive to the inclusion of various demographic control variables. This is unsurprising given that our samples are reasonably large and well-randomized. Second, we demonstrate that our main results are not attributable either to special features of particular experimental tasks such as the time horizon, or to special subgroups of subjects defined by initial levels of financial literacy, degree of responsiveness to variation in experimental stimuli, or degree to which a subject's implied rate of time preference is stable across simply framed tasks. Third, we adapt our analysis to allow for the possibility that subjects' valuations may be "fuzzy." Here we employ two distinct analytic strategies. One is to assume that "true" valuations are well-defined, and that the fuzziness reflects noisy elicitation, which could in principle mask improvements in welfare. The other strategy is to proceed according to the Bernheim-Rangel welfare framework, treating fuzzy valuations as implying normative ambiguity. Both strategies leave our qualitative conclusions unchanged.

7 Generalizability

Naturally, one must exercise caution when generalizing from any study that focuses on a single financial education intervention. Certainly, our analysis does not justify a broad inference that financial education programs fail to improve welfare. Different interventions may have different effects. Indeed, even the same intervention may produce dissimilar outcomes in different populations.

Even so, our analysis does have important general implications. First, it highlights the potential pitfalls of educational interventions that are brief and laden with motivational rhetoric. In these and other respects, the intervention we examine is typical of the programs offered to millions of workers through their employers.[46] Unfortunately, as we have seen, behavior may respond primarily to motivational rhetoric even when people appear to understand and internalize the substantive elements of instruction. By making the material engaging and memorable, educators may also render its behavioral effects indiscriminate, and consequently of questionable value. This conclusion directly challenges received wisdom and argues for a reexamination of the principles governing the design of adult financial education interventions.

Second, we have shown that one cannot count on conventional outcome measures to reliably detect these deficiencies. The intervention we consider improves measured financial literacy, increases the self-reported use of the desired decision strategy without reducing reliance on advice or other analytic methods, and on average counteracts a known decision bias. Even so, it fails to enhance the average quality of decision making in simple choice tasks that are designed to permit easy application of the targeted tools. Our analysis therefore underscores the importance of conducting explicit and rigorous examinations of welfare when evaluating particular educational interventions. Moreover, our notion of financial competence leads to practical welfare measures that address this need.

It is of course appropriate to ask whether the welfare effects measured through our method are generalizable beyond the diagnostic tasks from which they are derived. A skeptic might raise one of three issues.

First, improved performance in simple diagnostic tasks does not necessarily translate into better real-world decision making in the complex contexts educational interventions target, such as saving

[46]Another representative feature is that the intervention lacks opportunities for practice and feedback.

for retirement. But if an intervention does not even improve the quality of decision making in tasks to which the pertinent concepts are easily applied, the notion that it will do so in more complex real-world settings, except by chance, is far-fetched.[47] Moreover, if one finds improvements in the simplest diagnostic tasks, one can then deploy the same methods in a sequence of increasingly complicated tasks that mimic additional features of real-world problems.

Second, a failure to improve performance in diagnostic tasks might not generalize to real-world contexts with higher stakes. Plainly, the diagnostic stakes must be large enough to motivate subjects. In the context of our current experiment this is certainly the case, in that we recruit our subjects through an online labor market, and offer payments that are substantial in comparison to the wages for which they normally work. One cannot plausibly attribute the absence of welfare gains to insufficient stakes, inasmuch as the intervention significantly improves performance on a incentivized test of financial literacy even though the stakes are also small.

Third, a failure to improve performance in diagnostic tasks might not generalize to real-world contexts wherein people may be more likely to employ analytic tools or seek advice. We note, however, that roughly three-quarters of the US population reports making real financial decisions without assistance (Lusardi and Mitchell, 2011). Additionally, by design, we did not preclude subjects from seeking help. Indeed, because we conducted our experiment online, subjects had access to all resources available through the internet, and were given ample time to use them. Notably, the fraction of individuals reporting that they did not seek advice in our experiment (three quarters) matches experience in the field.

8 Conclusion

In this paper, we introduced a new method for measuring the quality of financial decision making built around the concept of *financial competence*. We used this notion to document the potential pitfalls of the types of brief rhetoric-laden interventions that are commonly used for adult financial education. We also demonstrated that conventional methods of evaluation do not reliably detect these deficiencies, thereby establishing the importance of including assessments of financial competence in evaluations of educational interventions.

We say that consumers are financially competent with respect to specific financial principles if they make equivalent choices from equivalent opportunity sets whenever an understanding of those principles would enable them to verify the equivalencies. To assess financial competence, we compare

[47] In principle, the loss function for decisions made in the field could be asymmetric, for instance with underestimation of compound interest more damaging than overestimation. In that case, the fact that we have designed diagnostic tasks with symmetric loss functions could cause us to underestimate the benefits of a measure that causes an indiscriminate increase in the valuations ascribed to interest-bearing assets. However, any improvement in welfare is then entirely fortuitous, and not the result of enhanced decision-making skill.

a consumer's decisions across equivalent complexly framed and simply framed valuation tasks. As a method of evaluating the quality of financial decision making, this new approach offers a number of significant advantages over conventional metrics: it is non-paternalistic, it imposes modest information requirements, it is simple, intuitive, and easily implemented; and it yields a quantitative measure of the quality of financial decision making which, under relatively modest assumptions, is formally interpretable in terms of consumer welfare.

The financial education intervention we study resembles typical employer-sponsored programs with respect to its brevity and emphasis on heuristic and motivational messages; subject to the constraints of brevity, it is ostensibly well-designed. Indeed, we find that it significantly improves measured financial literacy, and subjects report that they operationalize their improved knowledge when making choices. The intervention even eliminates exponential growth bias on average. However, financial competence does *not* improve. Further investigation reveals the explanation: behavior responds primarily to motivational rhetoric even when people appear to understand and internalize the substantive elements of instruction. While the rhetorical components make the material engaging and memorable, they also render its behavioral effects indiscriminate, and consequently of limited value.

Our main findings have potentially important implications for public policy. Potential strategies for addressing deficiencies in financial competence fall into three broad categories. The first is to devise educational methods that more effectively lead people to put pertinent knowledge into practice when making decisions, and to do so correctly. Given that brevity appears to be a design constraint for adult financial education, it is important to determine whether efficacy and brevity are compatible. In light of our analysis, we recommend exploring program designs that replace motivational rhetoric and simple prescriptive dicta with practical exercises that illustrate the application of the pertinent principles and that create opportunities for providing participants with practice and feedback. Rhetorical prods may be useful for the purpose of marketing educational programs and boosting participation, but counterproductive when incorporated into pedagogy. Our analysis has also shown that it is important to evaluate the success of alternative designs by assessing financial competence using the methods we have developed, rather than by administering simple tests of financial knowledge.

A second strategy is to deploy educational programs targeted at populations known to manifest particular biases in order to create countervailing biases. In effect, this amounts to accomplishing the right objective for the wrong reason. To illustrate, in the current study, we have found that the most beneficial intervention is actually the one with the least substance and the most rhetorical motivation. Presumably, we could enhance its aggregate benefit by limiting its deployment to subjects whose demographic characteristics and initial test scores indicate a high degree of susceptibility to exponential growth bias. This "targeted de-biasing" strategy is likely to prove challenging, however, because it seems likely that any success in balancing countervailing biases will be highly context-specific.

A third strategy is to simplify the framing of naturally occurring decision problems, either by developing and deploying better tools for visualizing opportunities and consequences, or by requiring suppliers of financial products to characterize them in simple terms. In principle this is a promising approach, but its effective implementation will require much additional research.

Having developed a framework for answering practical questions about financial competence, we envision many directions for subsequent research, some of which we are already pursuing. One important task is to extend our methods to other types of financial decisions such as insurance and portfolio allocation, involving concepts such as risk taking, inflation, and management fees. It is also important to study other populations, as well as other types of educational interventions, particularly ones that are used in practice. Accordingly, we anticipate using these methods to evaluate actual adult educational interventions in the workplace and other settings. Research on pedagogical design will, however, at least initially require extensive study of more narrowly focused interventions in the laboratory. Indeed, a focus on narrow educational interventions makes it easier to determine which pedagogical approaches work and which do not, and to develop a nuanced understanding of the mechanisms through which such interventions influence behavior. For these reasons, we have reservations concerning the call in Hastings, Madrian and Skimmyhorn (2013) for studies of "large scale interventions." The effective design of such interventions likely requires a much more comprehensive micro-level understanding of financial education than we currently possess. An initial focus on narrow small-scale interventions is, in our view, the best route to developing that understanding.

In principle, our methods could be used to evaluate other types of educational interventions that aim to provide people with a better understanding of their choice's consequences. Applications to problems involving health and nutrition are worth exploring.

References

Abeler, Johannes and Simon Jaeger, "Complex Tax Incentives," *American Economic Journal: Economic Policy*, 2014.

Afriat, Sidney N., "Efficiency Estimation of Production Functions," *International Economic Review*, 1972, *13* (3), 568–98.

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David Laibson, "The Age of Reason: Financial Decisions over the Life Cycle and Implications for Regulation," *Brookings Papers on Economic Activity*, 2009, *Fall*, 51–101.

Almenberg, Johan and Christer Gerdes, "Exponential growth bias and financial literacy," *Applied Economics Letters*, 2012, *19* (17).

Andersen, Steffen, Glenn W. Harrison, Mortel I. Lau, and E. Elisabet Rutstrom, "Elicitation using multiple price list formats," *Experimental Economics*, 2006, *9*, 383–405.

Arrow, Kenneth J., "Rational Choice Functions and Orderings," *Economica*, 1959, *26* (102), 121–127.

Aufenanger, Tobias, Friedemann Richter, and Matthias Wrede, "Measuring Decision-Making Ability in the Evaluation of Financial Literacy Education Programs," *Unpublished Manuscript*, 2016.

Austin, Rob and Winfield Evens, "2013 Trends & Experience in Defined Contribution Plans," *Aon Hewitt*, 2013.

Baltussen, Guido and Gerrit T. Post, "Irrational Diversification: An Examination of Individual Portfolio Choice," *Journal of Financial and Quantitative Analysis*, 2011, *5*, 1463 – 1491.

Bayer, Patrick J., B. Douglas Bernheim, and John Karl Scholz, "The Effects of Financial Education in the Workplace: Evidence from a Survey of Employers," *Economic Inquiry*, 2009, *47* (4), 605–624.

Bernheim, B. Douglas, "Behavioral Welfare Economics," *Journal of the European Economic Association*, 2009, *7* (2-3), 267–319.

_ , "The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics," *Journal of Benefit-Cost Analysis*, 2016, *7* (1), 12–68.

_ **and Antonio Rangel**, "Addiction and Cue-Triggered Decision Processes," *American Economic Review*, 2004, *94* (5), 1558–90.

— **and** — , "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *The Quarterly Journal of Economics*, 2009, *124* (1), 51–104.

— **and Daniel M. Garrett**, "The effects of financial education in the workplace: Evidence from a survey of households," *Journal of Public Economics*, 2003, *87*, 1487–1519.

— **,** — **, and Dean M. Maki**, "Education and saving: The long-term effects of high school financial curriculum mandates," *Journal of Public Economics*, 2001, *80*, 435–465.

Bertrand, Marianne and Adair Morse, "Information Disclosure, Cognitive Biases, and Payday Borrowing," *The Journal of Finance*, 2011, *LXVI* (6), 1865–93.

Brown, Alexandra, J Michael Collins, Maximilian Schmeiser, and Carly Urban, "State mandated financial education and the credit behavior of the young," *Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C., Finance and Economics Discussion Series*, 2014, *2014-68.*

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini, "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes," *Journal of Political Economy*, 2007, *115* (5), 707–47.

— **,** — **, and** — , "Measuring the Financial Sophistication of Households," *American Economic Review: Papers & Proceedings*, 2009, *99* (2), 393–398.

Caplin, Andrew and Daniel Martin, "A testable theory of imperfect perception," *The Economic Journal*, 2015, *125*, 184–202.

— **and Mark Dean**, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 2015, *105* (7), 2183–2203.

Carlin, Bruce I., Li Jiang, and Stephen A. Spiller, "Learning Millennial-Style," *Working Paper, Anderson School of Business, UCLA*, 2014.

Carpena, Fenella, Shawn Cole, Jeremy Shapiro, and Bilal Zia, "Unpacking the Causal Chain of Financial Literacy," *The World Bank Policy Research Working Paper*, 2011, *5798.*

Choi, James J., David Laibson, and Brigitte C. Madrian, "$100 Bills on the Sidewalk: Suboptimal Investment in 401(k) Plans," *Review of Economics and Statistics*, 2011, *93* (3), 748–763.

Choi, Syngjoo, Shachar Kariv, Wieland Mueller, and Dan Silverman, "Who is (More) Rational?," *American Economic Review*, 2014, *104* (6), 1518–1550.

Cole, Shawn and Gauri Kartini Shastry, "Is high school the right time to teach self-control? The effect of education on financial behavior.," *Unpublished Manuscript, Harvard University.*, 2012.

__ **, Thomas Sampson, and Bilal Zia**, "Prices or Knowledge? What Drives Demand for Financial Services in Emerging Markets?," *The Journal of Finance*, 2011, *66* (6), 1933–1967.

Collins, J.M., "The impacts of mandatory financial education: Evidence from a randomized field study.," *Working Paper, Center for Financial Security, University of Wisconsin-Madison*, 2010.

Council for Economic Education, "Financing your future (DVD)," `http://financingyourfuture.counciloreconed.org/` 2006.

Drexler, Alejandro, Greg Fischer, and Antoinette Schoar, "Keeping It Simple: Financial Literacy and Rules of Thumb," *American Economic Journal: Applied Economics*, 2014, *6* (2), 1–31.

Duflo, Esther and Emmanuel Saez, "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *Quarterly Journal of Economics*, 2003, *118* (3), 815–842.

Echenique, Frederico, Sangmok Lee, and Matthew Shum, "The Money Pump as a Measure of Revealed Preference Violations," *Journal of Political Economy*, 2011, *119* (6), 1201–1223.

Eisenstein, Eric M. and Stephen J. Hoch, "Intuitive Compounding: Framing, Temporal Perspective, and Expertise," *Unpublished Manuscript*, Dec 2007.

Enke, Benjamin and Florian Zimmermann, "Correlation neglect in belief formation," *Unpublished Manuscript*, 2013.

Ernst, Keith, John Farris, and Uriah King, "Quantifying the Economic Cost of Predatory Payday Lending," Technical Report, Center for Responsible Lending 2004.

Fernandes, Daniel, John G Lynch, and Richard G Netemeyer, "Financial literacy, financial education, and downstream financial behaviors," *Management Science*, 2014, *60* (8), 1861–1883.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue, "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, 2002, *40* (2), 351–401.

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

_ , Matthew R Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff, "The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings," *NBER working paper*, 2015.

Harrison, Glenn W, Morten Igel Lau, E Elisabet Rutström, and Melonie B Sullivan, "Eliciting risk and time preferences using field experiments: Some methodological issues," *Field experiments in economics*, 2005, *10*, 125–218.

Hastings, Justine S. and Lydia Tejeda-Ashton, "Financial Literacy, Information, and Demand Elasticity: Survey and Experimental Evidence from Mexico," *NBER Working Paper*, 2008, *14538.*

_ , Brigitte C. Madrian, and William L. Skimmyhorn, "Financial Literacy, Financial Education, and Economic Outcomes," *Annual Review of Economics*, 2013, *5*, 347–373.

Heinberg, Aileen, Angela A. Hung, Arie Kapteyn, Annamaria Lusardi, Anya Savikhin Samek, and Joanne K. Yoong, "Five steps to planning success. Experimental Evidence from U.S. Households," *Oxford Review of Economic Policy*, 2014, *30* (4), 697–724.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *American Economic Review*, 2002, *92(5)*, 1644 – 1655.

Horton, John J., David G. Rand, and Richard J. Zeckhauser, "The online laboratory: Conducting experiments in a real labor market," *Experimental Economics*, 2011, *14*, 399–425.

Jump$tart Coalition for Personal Financial Literacy, "Financial literacy shows slight improvement among nation's high school students," *Report*, 2006, *Washington, D.C.*

Kalaci, Kenan and Marta Serra-Garcia, "Complexity and Biases," *Experimental Economics*, 2016, *1*, 31–50.

Kline, Paul, *Handbook of Psychological Testing*, 2 ed., London and New York: Routledge, 1999.

Kunda, Ziva, "The case for motivated reasoning.," *Psychological bulletin*, 1990, *108* (3), 480.

Levy, Matthew R. and Joshua Tasoff, "Exponential-Growth Bias and Overconfidence," *unpublished manuscript*, 2015.

_ and _ , "Exponential Growth Bias and Life Cycle Consumption," *Journal of the European Economics Association*, 2016, *14* (3), 545–583.

Lührmann, Melanie, Marta Serra-Garcia, and Joachim Winter, "Teaching teenagers in finance: Does it work?," *Munich Discussion Paper*, 2012, *24.*

_ , _ , and _ , "The Impact of Financial Education on Adolescents' Intertemporal Choices," *Unpublished Manuscript, University of Munich*, 2014.

Lusardi, Annamaria, "Americans' Financial Capability," *NBER Working Paper*, 2011, *17103.*

_ and Olivia Mitchell, "How ordinary consumers make complex economic decisions: financial literacy and retirement readyness," *NBER Working Paper*, 2009, *15350.*

_ and _ , "Financial Literacy and Planning: Implications for Retirement Well-being," in Annamaria Lusardi and Olivia Mitchell, eds., *Financial Literacy. Implications for Retirement Security and the Financial Marketplace*, Oxford University Press, 2011, pp. 17–39.

_ and _ , "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, 2014, *52* (1), 5–44.

_ , Anya Savikhin Samek, Arie Kapteyn, Lewis Glinert, Angela Hung, and Aileen Heinberg, "Visual Tools and Narratives: New Ways to Improve Financial Literacy," *Journal of Pension Economics and Finance*, forthcoming.

Malkiel, Burt G. and Charles D. Ellis, *The Elements of Investing. Easy Lessons for Every Investor.*, New Jersey: Wiley, 2013.

Mandell, Lewis, "The Financial Literacy of Young American Adults: Results of the 2008 National Jump$tart Coalition Survey of High School Seniors and College Students," *Jump$tart Coalition*, 2009, *Washington, D.C.*

_ and Linda Schmid Klein, "The Impact of Financial Literacy Education on Subsequent Financial Behavior," *Journal of Financial Counseling and Planning*, 2009, *20* (1), 15–24.

Mason, Winter and Siddarth Suri, "Conducting behavioral research on Amazon's Mechanical Turk," *Behavioral Research*, 2012, *44*, 1–23.

McKenzie, Craig R. M. and Michael J. Liersch, "Misunderstanding Savings Growth: Implications for Retirement Savings Behavior," *Journal of Marketing Research*, 2011.

O'Donoghue, Ted and Matthew Rabin, "Doing It Now or Later," *The American Economic Review*, 1999, *89* (1), 103–124.

Peysakhovich, Alexander, Martin A. Nowak, and David G. Rand, "Humans display a 'cooperative phenotype' that is domain general and temporally stable," *Nature Communications*, 2014, *5.*

Servon, L.J. and R. Kaestner, "Consumer financial literacy and the impact of online banking on the financial behavior of lower-income bank customers," *Journal of Consumer Affairs*, 2008, *42*, 271–305.

Skimmyhorn, William L., "Essays in behavioral household finance." PhD dissertation, Harvard Kennedy School, Cambridge, MA 2012.

_ , "Assessing Financial Education: Promising Evidence From Boot Camp," *USMA Working Paper*, 2015.

Song, Changcheng, "Financial Illiteracy and Pension Contributions: A Field Experiment on Compound Interest in China," *Unpublished Manuscript*, March 2015.

Stango, Victor and Jonathan Zinman, "Exponential Growth Bias and Household Finance," *Journal of Finance*, 2009, *64* (6), 2807–2849.

Taubinsky, Dmitry and Alexander Robert Rees-Jones, "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *Unpublished Manuscript*, 2015.

van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie, "Financial literacy and stock market participation," *Journal of Financial Economics*, 2011, *101*, 449–472.

Wagenaar, William M. and Sabato D. Sagaria, "Misperception of Exponential Growth," *Perception and Psychology*, 1975, *18* (6), 416–422.

Walstad, William B., Ken Rebeck, and Richard A. Macdonald, "The Effects of Financial Education on the Financial Knowledge of High School Students," *The Journal of Consumer Affairs*, 2010, *44* (2), 336–357.

This and other Global Financial Literacy Excellence Center working papers and publications are available online at
www.gflec.org

GFLEC
GLOBAL FINANCIAL LITERACY
EXCELLENCE CENTER

Global Financial Literacy Excellence Center
The George Washington University
School of Business
Duquès Hall, Suite 450
2201 G Street NW
Washington, DC 20052

P: 202-994-7148
F: 202-994-8289
E: gflec@gwu.edu