NBER WORKING PAPER SERIES

THE EFFECT OF FINANCIAL EDUCATION ON THE QUALITY OF DECISION MAKING

Sandro Ambuehl
B. Douglas Bernheim
Annamaria Lusardi

Working Paper 20618
http://www.nber.org/papers/w20618

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2014

Previously circulated as "Financial Education, Financial Competence, and Consumer Welfare." We thank Charles Sprenger, Steven Sheffrin, 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, and the Murphy Institute's conference on Expanding the Frontiers in Behavioral Public Economics 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Effect of Financial Education on the Quality of Decision Making
Sandro Ambuehl, B. Douglas Bernheim, and Annamaria Lusardi
NBER Working Paper No. 20618
October 2014, Revised July 2014
JEL No. D14,D91,I21

ABSTRACT

We introduce the concept of financial competence, a measure of how closely individuals' choices align with those they would make if they properly understood their opportunity sets. The concept is firmly rooted in the principles of choice-based behavioral welfare analysis and avoids the types of paternalistic judgments that pervade policy discussions. We document the importance of assessing financial competence by demonstrating experimentally that an educational intervention can appear highly successful according to conventional outcome measures while failing to improve the quality of financial decision making. We trace the mechanisms behind these seemingly divergent findings.

Sandro Ambuehl
Department of Economics
Stanford University
Stanford, CA 94305
sambuehl@stanford.edu

B. Douglas Bernheim
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER
bernheim@stanford.edu

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

An Online Appendix is available at:
http://www.nber.org/data-appendix/w20618

"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 serious questions about the general quality of financial decision making. Financial education aims to improve decision making by helping consumers acquire the basic knowledge and skills they need to understand the choices they face. A large and growing literature examines the effects of financial education on both financial literacy (as measured by test scores) and financial choices (such as saving).[1] Unfortunately, that literature sheds little objective light on the critical question of whether the behavioral effects of financial education are helpful or harmful. Discussions of these issues are typically colored by paternalistic judgments (for example, that people are better off with high saving and balanced portfolios) and/or strong preconceptions (for example, that a better understanding of choice options necessarily promotes better decisions). Yet it is also possible that financial education alters behavior through mechanisms that involve indoctrination, exhortation, deference to authority, social pressure, or psychological anchors, in which case it may induce people to act contrary to the preferences they themselves would reveal through choices if they properly understood the consequences of their actions. Because the existing literature does not distinguish between these mechanisms, it ultimately has little to say about the welfare consequences of financial education. (See Section 9 for a more detailed discussion of existing research that bears on this issue.)

This paper makes three main contributions. First, we introduce a new approach to evaluating the quality of financial decision making. One cannot make such evaluations within the revealed preference paradigm, because all choices tautologically serve the objectives they reveal. Stepping outside of that paradigm frees us of the tautology, but also necessitates the adoption of other organizing principles. We proceed from the sensible premise that good decision making requires a proper understanding of the relationships between actions and consequences. A consumer who both possesses and acts upon such an understanding will exhibit *consistency* across equivalent representations of the same decision problem, irrespective of her preferences. Accordingly, we say that a consumer is *financially competent*

[1] See Lusardi and Mitchell (2014) and Hastings et al. (2013) for recent comprehensive reviews of this literature.

with respect to specific financial principles if she makes equivalent choices from equivalent opportunity sets whenever an understanding of those principles would enable her to verify the equivalencies.[2]

More concretely, our approach involves comparisons of a consumer's decisions across equivalent *complexly framed* and *simply framed valuation tasks*. Each such task elicits the consumer's current dollar equivalent value for the cash flows associated with some standard financial instrument. For a complexly framed task, the consumer receives a complete description of the instrument; for the equivalent simply framed task, she receives a transparent description of the associated cash flows. If the consumer fully understands the pertinent financial principles governing the instrument and acts on that understanding, the valuations should be identical. When the valuations differ, the magnitude of the discrepancy reflects the consumer's degree of financial competence. To the extent financial education induces consumers to incorporate knowledge of applicable financial principles into decision making, it should bring equivalent simply and complexly framed valuations into closer alignment, thereby improving measured competence.

Our approach to measuring the quality of financial decision making offers several advantages over more conventional alternatives. 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, as we explain in details below, the approach provides a quantitative measure of the quality of financial decision making which, under relatively modest assumptions, is formally interpretable in terms of consumer welfare, and thus amenable to cost-benefit analysis. Third, it imposes modest information requirements. By comparing a consumer's choices for equivalent tasks, we avoid the need for theoretical explanations and detailed preference-based models of decision making. Fourth, because the evaluation is performed subject-by-subject, it explicitly and flexibly accounts for population heterogeneity. This is important because different people likely start out with different levels of knowledge and even opposite biases, so that one would expect the effects of financial education to vary widely. Yet most previous studies ignore such heterogeneity, and those that allow for it do so in a comparatively crude manner.[3]

Our second main contribution is to document empirically the importance of assessing and analyzing financial competence, rather than relying exclusively on conventional outcome measures. We demonstrate that an educational intervention with features typical of workplace financial education programs (as detailed in the next paragraph) can appear highly successful according to those conventional measures while failing to improve the quality of financial decision making.

[2]The literature uses the term *financial capability* to signify the quality of financial decision making. Because that existing term lacks a precise definition, we avoid it. Our notion of *financial competence* is, of course, related, but it has a specific meaning.

[3]For example, Bernheim et al. (2001) found that exposure to financial education in high school increased saving among people who described their parents as "not frugal" but not among those with "frugal" parents.

One important feature of workplace financial education programs is that they are *brief*[4]. Brevity is effectively a design constraint: more thorough approaches are costly and time-consuming, which makes them unappealing to both employers and workers. To compensate for brevity, these programs generally have a second important feature; they tend to focus on simple, memorable, and potentially useful heuristics accompanied by highly motivating messages. Various studies have found that households change their financial choices in response to these types of interventions (e.g. Bayer et al. (2009), Duflo and Saez (2003)); whether they are better or worse off as a result is unclear.

We use our approach to evaluate the quality of decision making in a setting that requires an understanding of compound interest, one of the fundamental concepts in personal finance. This application is of considerable practical interest because it is well-established that people tend to suffer from *exponential growth bias*, the tendency to underestimate the growth of an investment when interest is compounded (see Wagenaar and Sagaria (1975), Eisenstein and Hoch (2007), Stango and Zinman (2009), Almenberg and Gerdes (2012), Levy and Tasoff (ming)).

In our experiment, some subjects received financial education pertaining to compound interest, while others did not.[5] The educational module resembles typical employer-sponsored interventions with respect to its brevity and emphasis on heuristic and motivational messages; subject to the constraints of brevity, it is ostensibly well-designed. Treated and untreated subjects performed ten simply framed valuation tasks, along with ten equivalent complexly framed valuation tasks, all with real monetary consequences. For example, if the complexly framed task elicited the value of X for which a subject is indifferent between \$$X$ immediately and \$10 invested for 36 days at a return of 2% per day, the associated simply framed task elicited the value of Y for which she is indifferent between \$$Y$ immediately and \$20 in 36 days. Finally, subjects took an incentivized test on compound interest, and answered questions concerning their decision-making strategies.

Our main findings are as follows. The intervention substantially improves subjects' knowledge and conceptual understanding of compound interest (financial literacy), as measured by their incentivized test scores. Moreover, subjects report that they operationalize the newly gained knowledge in their consequential decisions. The fraction of subjects who say they make decisions based on numerical calculations rises sharply, and the fraction who say they seek external help does not decline, which suggests that the new tools do not simply crowd out other potentially reliable approaches to the valuation tasks. The intervention also significantly increases valuations for complexly framed tasks, which is precisely what one would hope to find if it effectively counteracts exponential growth bias.

[4]Fernandes et al. (2014) report 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.

[5]As detailed below, subjects were recruited from the crowdsourcing platform Amazon Mechanical Turk, and were generally of working-age (20s and 30s), with lower-than-average incomes but higher-than-average education. Their *ex ante* understanding of compound interest was generally poor.

Thus, an analysis based on conventional outcome metrics such as financial literacy, self-reported decision strategies, and/or directional effects on choice would conclude that – despite its brevity – the intervention was highly successful and presumably welfare-enhancing. Yet our analysis of financial competence (the magnitude of the typical gap between X and Y) paints a much different picture. Surprisingly, the full intervention has *no effect whatsoever* on the average quality of financial decision making.[6]

Our third main contribution is to explore the mechanisms that produce apparently divergent effects on conventional outcome metrics and financial competence. Because our analysis is confined to a single limited educational intervention, we caution against generalizing from our conclusions concerning these mechanisms. Still, we regard this portion of our investigation as an important step toward understanding the relationships between knowledge, motivation, and choice, and hence designing more effective and beneficial educational interventions, particularly subject to the constraint of brevity.

As is typical of financial education interventions, ours includes both substantive material and motivational rhetoric, such as atmospheric declarations about the power of compounding and exhortations to save. To investigate mechanisms, we fielded two additional interventions, one limited to the substantive portions of the original intervention, the other limited almost exclusively to the rhetoric. A comparison of results for our three interventions reveals that the effects on measured financial literacy and self-reported decision strategies are due almost entirely to the inclusion of substantive material, as one would hope. In contrast, the effects on choices in valuation tasks are almost entirely attributable to rhetoric, not to substance: the substance-only intervention does not have a significant impact on average valuations for complexly framed choices, while the effect of the rhetoric-only intervention is statistically indistinguishable from that of the full intervention.

A closer examination of our results also reveals that the effect of the full treatment on choices in complexly framed valuation tasks bears little or no relation to the subject's initial degree of exponential growth bias. Ideally, the intervention would increase valuations among those undervaluing complexly framed options, and decrease valuations by those overvaluing those options. Yet on average, it increases valuations in the complexly framed problems across the board (while leaving valuations in the simply framed problems unchanged). This is precisely what one would expect given that the effect is apparently attributable to rhetoric rather than substance.

Several additional findings concerning mechanisms merit emphasis. First, while the substance-only treatment causes subjects to make complexly framed choices more slowly (which suggests more careful deliberation), this effect disappears with the addition of rhetoric. We conjecture that rhetoric

[6] A vast literature on transfer learning has yet to yield general conclusions (see Barnett and Ceci (2002) for a review). As a whole, this literature neither predicts nor contradicts our findings.

accelerates these choices by inducing subjects to substitute a less cognitively demand heuristic for rigorous calculations. If that conjecture is correct, it raises the more general possibility that the use of motivational rhetoric may defeat the purpose of substantive instruction. Second, the success of our substantive interventions at improving measured financial competence in any given problem is unrelated to the difficulty of applying the concepts in that problem. Thus it appears that our results are not attributable to the complexity of the tasks or the limited nature of the intervention. Third, all of our interventions reduce reliance on simple interest calculations. It therefore appears that intellectual stubborness is not the governing consideration. Rather, the interventions apparently migrate subjects to other similarly inappropriate decision strategies.

Taken together, these findings call into question the validity of efforts to evaluate the benefits of educational interventions through analyses that are confined to effects on financial literacy, directional changes in behavior, and/or changes in self-reported decision strategies. Our results also highlight the pitfalls of policy agendas that specifically target only the aforementioned objectives. At the same time, we offer a conceptually rigorous and practical alternative.

The remainder of the paper is organized as follows. Section 2 introduces the concept of financial competence, defines it precisely, discusses its measurement, and provides a formal welfare interpretation. Section 3 describes our experiment, and section 4 discusses its implementation. Sections 5, 6, and 7 analyze the effects of the treatments on test scores, average choices, and consumer welfare, respectively. Section 8 analyzes the channels through which the interventions affect behavior. Section 9 relates our research to the existing literature. Section 10 discusses the policy implications of our research and concludes.

2 The Definition and Measurement of Financial Competence

One of our main objectives is to devise a general framework for evaluating financial competence in a wide range of decision-making contexts, and not simply an *ad hoc* approach to the particular context studied later in this paper. Accordingly, we begin with a general definition of financial competence that identifies at an abstract conceptual level the precise feature of behavior we wish to assess. Moving from this abstract definition to any dataset that one is likely to possess presents challenges and requires us to make some assumptions, which we identify explicitly. We then turn to the critical question of how one might measure the extent to which actual decisions depart from the ideal of perfect competence. We propose an intuitively appealing measure, and show that it admits a precise welfare interpretation under relatively weak conditions.

2.1 A general definition of financial competence

The term "financial competence," as we use it, references a consumer's ability to make good financial decisions. From a non-paternalistic perspective, whether a given decision is good or bad plainly depends on the decision maker's objectives. Unfortunately, we cannot observe those objectives directly.

If we adopt the standard view that choices reveal objectives, then all choices are good by definition, and competence is tautological. A more nuanced view holds that high-quality financial decision making requires an understanding of the relationships between choices and outcomes. It is commonly assumed that financial literacy, defined as a mastery of the conceptual principles behind those relationships, provides that understanding. If that is the case, then one can equate financial competence with financial literacy. We reject that approach on the grounds that consumers either may not try to put principles into practice, or may deploy those principles incorrectly.

Our approach to defining financial competence is predicated on three simple observations. First, financial decisions generally involve choices among what we will call *consumption instruments*, rather than consumption bundles. Consumption instruments are derivative goods that are valued only because they provide the means to secure bundles of intrinsically valued goods. For example, consumers obtain future (as well as state-contingent) consumption by making decisions about saving and investments. Second, there are typically many ways to assemble menus of instruments that yield the same consumption opportunities, and hence constitute equivalent decision problems. We will elaborate on this principle shortly. Third, a consumer who acts on a proper understanding of the relationships between actions and consequences should exhibit *consistency* across equivalent representations of the same decision problem, irrespective of her preferences. Thus, financial competence (as we define it) entails equivalent choices from equivalent decision problems. Significantly, note that under this definition, we can assess competence without determining the consumer's objectives.

To provide a formal and general definition of financial competence, we must introduce some notation. In standard consumer theory, we think of the consumer as attaching intrinsic value to elements of a consumption set $\mathcal{C}$, and as making choices from an opportunity set $C \in \mathcal{C}$. Here we are concerned instead with choices among consumption instruments. We will use $\mathcal{I}$ to denote the set of all conceivable instrumental alternatives that might be available to a decision maker at a particular point in time, $I \subseteq \mathcal{I}$ to denote an opportunity set,[7] and $2^{\mathcal{I}}$ to denote the set of possible opportunity sets. The consumer obtains an element of $\mathcal{C}$ through a process that begins with the selection of some $i \in I$, but typically does not end with that choice. Instead, opting for a given alternative usually determines subsequent opportunities and necessitates additional decisions. For instance, investments generate

[7] $\mathcal{I}$ is defined to include all trivial instruments, including the elements of $\mathcal{C}$ itself.

returns, which consumers can then choose to spend or reinvest. Thus, each alternative i maps to a *menu* of feasible consumption bundles, rather than to a single element of $\mathcal{C}$. We will use $\Gamma : \mathcal{I} \rightrightarrows \mathcal{C}$ to denote the correspondence identifying the consumption opportunities that remain available to the consumer after choosing a particular instrumental alternative.

Two opportunity sets, I and J, are *equivalent* iff $\bigcup_{i \in I} \Gamma(i) = \bigcup_{j \in J} \Gamma(j)$.[8] In words, equivalent opportunity sets make it possible for the consumer to obtain the same set of consumption bundles. For a given consumer, let $\gamma : 2^{\mathcal{I}} \rightrightarrows \mathcal{C}$ denote the choice correspondence specifying the consumption bundles she is willing to select through the process that begins with a choice of an instrument from some opportunity set $I \in 2^{\mathcal{I}}$. We define financial competence as the ability to act on the recognition that two decision problems are equivalent by ultimately selecting the same consumption bundle in each of them.

Definition. A consumer is *financially competent* if, for any pair of equivalent opportunity sets, I and J, he selects the same consumption bundle(s), $\gamma(I) = \gamma(J)$.

Notice that, by focusing on an aspect of internal consistency, we avoid the need to either make arbitrary assumptions about the nature of "true preferences," or justify comparisons to external paternalistic benchmarks. Even so, our approach is not assumption-free: to describe the set $\mathcal{C}$, one must take a stand on the aspects of experience that intrinsically matter to consumers.[9] Only then does the condition $\bigcup_{i \in I} \Gamma(i) = \bigcup_{j \in J} \Gamma(j)$ ensure that a consumer who fully understands the relationships between choices and consequences will view I and J as interchangeable.

An attractive feature of this definition is that it allows one to speak of domain-specific competencies: a consumer is financially competent with respect to specific financial principles if she makes equivalent choices from equivalent opportunity sets whenever an understanding of those principles would enable her to verify the equivalencies. Accordingly, when evaluating any particular educational intervention, one can focus on competencies within domains related to the educational content.

There are, of course, other notions of internal consistency such as WARP and GARP, but these are less well-suited to the task of assessing financial education interventions than our approach.[10]

[8] As a special case, we say that two instruments, i and j, are equivalent if $\Gamma(i) = \Gamma(j)$.

[9] It is worth emphasizing that without an assumption of this type, it would be impossible to evaluate the success or failure of financial education. For example, if we allow for the possibility that consumers care intrinsically about financial education and regard it as a potential complement or substitute for other goods, there is no way to distinguish between changes in consumption that reflect changes in preferences rather than changes in knowledge.

[10] We do not define financial competence in terms of conformance with choice axioms for the following reasons. First, financial education does not target such conformance 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). Second, a consumer who misunderstands a financial concept in a consistent manner may nevertheless respect such axioms. For example, one who incorrectly believes that instrument i will ultimately lead to a better consumption bundle than instrument j, perhaps because she uses the simple interest fomula to assess compound interest, will choose i over j, and will never choose j when i is available, thereby satisfying WARP (at least with respect to the choice of instruments). Third, our approach more readily yields measures of non-conformance that are interpretable as welfare losses. (To be clear, there are measures of non-conformance with GARP, such as the Afriat

2.2 Overcoming practical challenges

Moving from our abstract definition to any dataset one is likely to possess presents at least two practical challenges. First, if we interpret $\mathcal{C}$ as a set of life-cycle consumption bundles, the consumption-opportunities mapping Γ is complex. Determining whether two opportunity sets are equivalent is therefore potentially difficult, and could require many additional assumptions. Second, most datasets provide us with information about selected instruments (e.g., saving and investments) rather than lifetime consumption trajectories.

Our solution to the first challenge is to define *intermediate outcomes* as time-dated and state-contingent income profiles, and to call two instruments *strictly equivalent* if they lead to the same intermediate outcomes.[11] For example, two instruments that each deliver \$100 precisely one year from now in certain states of nature, and \$0 in all other states, are strictly equivalent. Significantly, if two instruments are strictly equivalent (lead to the same intermediate outcomes), then they are equivalent (potentially lead to the same sets of available consumption bundles). Thus, one can establish the equivalence of certain pairs of instruments without exhaustively enumerating all of the possibilities for spending income.

We address the second challenge by examining whether a consumer chooses strictly equivalent instruments from pairs of strictly equivalent equivalent opportunity sets. As a general matter, this condition neither implies, nor is implied by, our notion of financial competence (i.e., choosing the same consumption bundle from equivalent opportunity sets). However, the two can differ only if the consumer would eventually end up with different consumption bundles after choosing strictly equivalent instruments. For that eventuality to occur, the consumer would have to exhibit *persistent framing effects*, by which we mean that the framing of the original instrument continues to influence subsequent choices even after the intermediate outcome is realized. To justify our approach, we must therefore assume away any persistent framing effects.

2.3 Measuring financial competence in valuation tasks

Intuitively, we can assess a consumer's *degree* of financial competence by measuring the distance between choices made in strictly equivalent decision problems. Our approach involves conducting such analyses in the context of *valuation tasks*. A valuation task establishes the value $V(i)$ for which the consumer exhibits indifference between option i and \$$V(i)$ immediately (in conjunction with the status quo). In experimental settings, valuations are usually elicited through a collection of binary

(1972) critical cost efficiency index; see, e.g., Choi et al. (2014) for a related application. Morover, 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.)

[11] We say that two opportunity sets, I and J, are strictly equivalent iff for all $i \in I$ there is a strictly equivalent $j \in J$, and for all $j \in J$ there is a strictly equivalent $i \in I$.

choices (a *multiple price list*). Each opportunity set is of the form $I = \{i, d\}$, where i is an instrumental option and d represents an immediate dollar payment. We say that a consumer values i at $\$V(i)$ iff she chooses i when $d < V(i)$, and d when $d > V(i)$. Notice that the assessment of $V(i)$ necessarily involves a collection of choices (specifically, a range of values of d), rather than a single choice.

For our purposes, valuation tasks have several important virtues relative to other types of decision problems. First, and perhaps most important for our current purposes, valuations naturally lend themselves to easily interpreted notions of distance. Thus, if the valuations for two equivalent instrumental options, i and j, are (respectively) $V(i)$ and $V(j)$, then quantities such as $\big(V(i) - V(j)\big)^2$ and $\big|V(i) - V(j)\big|$ are intuitively appealing measures of the degree of financial competence. Moreover, as we explain in the next subsection, for appropriately selected options, these measures have precise welfare interpretations. Second, the component choices are extremely simple. Simplicity reduces the risk that differences in choices between equivalent decision problems may be attributable to factors other than limited comprehension of the concepts that financial education seeks to convey, such as a failure to notice or focus on a given option in a particular complex setting.[12] Third, while valuation tasks may at first seem somewhat artificial compared to real-world financial decisions, they are the building blocks for all other choices: if a consumer has coherent preferences over a set of options for which her valuations have been assessed, then one can infer the choices she would make for all opportunity sets consisting of those items.[13]

2.4 Welfare interpretation

The formal rationale for gauging financial competence based on the differences between the valuations of equivalent options rests on the principles of behavioral welfare economics, as articulated by Bernheim and Rangel (2009). In their framework, normative criteria are derived from choice patterns within the set of decision problems deemed *welfare-relevant*. Welfare-relevant choices may be internally consistent, in which case the approach amounts to revealed preference on a limited domain, or internally inconsistent, in which case the resulting criterion admits a degree of ambiguity.[14] Excluding a decision problem from the welfare-relevant set is warranted when the consumer's choice is predicated on a demonstrably incorrect understanding of available opportunities, in which case he is said to suffer from *characterization failure*.[15] Significantly, one can use this approach to measure the welfare loss associated with such failures.

[12] The ability to make good decisions when confronted with many options is another dimension of financial competence, but it is not typically the focus of financial education, and so is extraneous for our purposes.

[13] The literature documents cases in which valuations are imperfectly aligned with other choices (e.g. Grether and Plott (1979), Tversky et al. (1990)), but the settings are rather special. See Bordalo et al. (2012) for a discussion of the types of settings that tend to induce such misalignment.

[14] Formally, choices are internally consistent if they satisfy WARP, as defined by Arrow (1959).

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

We seek to apply this framework in a setting where the object is to evaluate the welfare consequences of limited financial competence within a domain involving specific conceptual principles. Accordingly, we first identify a class of instruments, $\mathcal{I}_C$ (such as fixed-term investments), that lead to consumption bundles through intermediate outcomes (such as deterministic or state-contingent income profiles), where knowledge of the specified principles enables one to derive the intermediate outcomes associated with each instrument. We then define a second class of instruments, $\mathcal{I}_S$, consisting of the intermediate outcomes themselves. We say that decision problems involving elements of $\mathcal{I}_C$ are *complexly framed*, while those involving $\mathcal{I}_S$ are *simply framed*.

For each complexly framed instrumental option $i_C \in \mathcal{I}_C$, there is a strictly equivalent simply framed option $i_S \in \mathcal{I}_S$. We will assume that there are no persistent framing effects, so that $\gamma(i_C) = \gamma(i_S) \equiv c$. In that case, if the consumer understands the consequences of her choices, we should observe $V(i_C) = V(i_S)$. If instead we find that valuations differ ($V(i_C) \neq V(i_S)$), then she presumably suffers from characterization failure. Assuming the evidence shows that she fails to understand an economic principle governing the mapping from complexly framed instruments to simply framed intermediate outcomes, then it is natural to infer that characterization failure is present in (at least) the choices involving complexly framed options.[16]

As a first case, suppose that the source of the characterization failure has been identified, and that it pertains *only* to the consequences of selecting the complexly framed instrument i_C. There is then a defensible rationale for excluding all the complexly framed decision problems from the welfare-relevant set, and (absent other evidence of characterization failure) retaining the simply framed ones. One can then use the latter to evaluate welfare losses in the former.

In this setting, $\big|V(i_C) - V(i_S)\big|$ represents the largest possible welfare loss the consumer may suffer as a result of characterization failure when making choices from sets of the form $\{i_S, d\}$. To illustrate, suppose first that $V(i_S) < V(i_C)$. If $d \leq V(i_S)$ or $d \geq V(i_C)$, there is no welfare loss associated with characterization failure, because the consumer makes the same choice from $\{i_S, d\}$ and $\{i_C, d\}$.[17] Mistakes occur when $V(i_S) < d < V(i_C)$. In this case, the consumer chooses i_C over \$$d$, even though she would willingly exchange $\gamma(i_C)$ for \$$d$, given a full understanding of the consequences. If she started out with her best option, \$$d$, she would be willing to give up \$$\big(d - V(i_S)\big)$ to avoid switching to $\gamma(i_C)$. Hence, \$$\big(d - V(i_S)\big)$ is the equivalent variation associated with the swap: it measures the dollar loss the consumer regards as equivalent to suffering the consequences of characterization failure. The loss is greatest when $d = V(i_C)$.

[16] To be clear, there are other possible inferences; see the discussion below.

[17] Technically, in the special case where $d = V(i_C)$, he definitely selects i_S from $\{i_S, d\}$ and is willing to select i_C from $\{i_C, d\}$.

Next suppose that $V(i_C) < V(i_S)$. As above, mistakes occur only when $V(i_C) < d < V(i_S)$. In this case, the consumer chooses \$$d$ over i_C, even though she would willingly exchange \$$d$ for $\gamma(i_C)$, given a full understanding of the consequences. If she started out with her best option, $\gamma(i_C)$, she would require $\$\big(V(i_C) - d\big)$ as compensation for switching to \$$d$. If income effects are negligible over the relevant range, compensating and equivalent variation coincide; $\$\big(d - V(i_C)\big)$ then measures the dollar loss the consumer regards as equivalent to suffering the consequences of characterization failure.[18] The loss is again greatest when $d = V(i_C)$. (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, but we leave that topic for a subsequent paper.)

Of course, for most values of d, the largest possible welfare loss overstates the actual loss. Another possibility is to compute the consumer's average (or expected) loss. Naturally, the expected loss depends upon the process that generates the consumer's opportunities. Our strategy is to calculate it in light of the process that actually generated the consumer's observed choices. 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 $\big|V(i_C) - V(i_S)\big|$, and the expected loss conditional upon suffering one is $\big|V(i_C) - V(i_S)\big|/2$; thus, the expected loss is proportional to $\big(V(i_C) - V(i_S)\big)^2$.[19]

The expressions for maximum and expected welfare losses, $\big|V(i_C) - V(i_S)\big|/2$ and $\big(V(i_C) - V(i_S)\big)^2$, are symmetric in i_C and i_S. To understand the importance of this observation, imagine that two economists disagree as to which choices are welfare relevant. One reasons as above, and concludes that valuation of the complexly framed instrument involves characterization failure. The other maintains that the complexly framed instrument has a form that is familiar to the consumer, while the simply framed one does not. He then asserts that the consumer has likely learned the relationship between complexly framed instruments and their consequences from experience, even if she does not understand the economic principles governing that relationship. Rather, she may invoke mistaken versions of those principles when "translating" a simply framed alternative into its more familiar complex form, in which case characterization failure plagues the valuation of the simply framed choice, not the

[18] 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.

[19] A possible objection to this observation is that, in our experiment, observed behavior ought to be invariant with respect to the distribution of choice problems (i.e., values of d); therefore, evaluating this behavior based on the uniform distribution is arbitrary. Fortunately, for any well-behaved distribution of d, the expected welfare loss is, to a good approximation, proportional to $\big(V(i_C) - V(i_S)\big)^2$. Formally, under the assumption that the CDF governing the distribution of d is twice differentiable, if we fix a value of $V(i_S)$ and take a second-order Taylor expansion of the expected welfare loss as a function of $V(i_C)$ in a neighborhood of $V(i_S)$, we obtain $\frac{\pi(V(i_S))}{2}\big(V(i_C) - V(i_S)\big)^2$, where π is the density function.

complexly framed one. Critically, there is no need to resolve this disagreement,[20] because swapping the roles of options i_C and i_S leave the formulas for maximum and expected welfare losses unchanged; both economists would reach the same conclusions concerning welfare.

The following is a summary of the assumptions under which our measure of financial competence admits a formal welfare interpretation.

Assumptions.

(i) The consumer does not intrinsically care about distinctions between two instruments that lead to the same intermediate outcomes.

(ii) There are no persistent framing effects. Two instruments that lead to the same intermediate outcomes lead to the same consumption choices.

(iii) Differences between the choices made in simply and complexly framed decision problems are attributable to characterization failure in one of the two settings.

A possible objection to the third assumption is that the processes relating most instrumental options to consumers' ultimate consumption bundles involve complicated economic principles; hence, characterization failure may be present in both simply and complexly framed decision problems. Yet even then, our welfare measures have meaningful interpretations under a weak alternative assumption: when the consumer evaluates one of the two classes of instruments (either the simply framed or complexly framed ones), she first translates them into elements of the other class. For concreteness, suppose she translates elements of $\mathcal{I}_C$ (complexly framed options) into elements of $\mathcal{I}_S$ (simply framed intermediate outcomes). A failure to understand the relationship between them implies that choice problems involving options in $\mathcal{I}_C$ induce characterization failure and should be excluded from the welfare-relevant set. The potential difficulty is that, if consumers also misunderstand the relationship between intermediate outcomes and consumption, choices involving elements of $\mathcal{I}_S$ are also suspect. Still, by retaining those choices in the welfare-relevant set, we can measure the loss the consumer would experience if her understanding of the relationship between intermediate outcomes and consumption bundles were correct. While that is not necessarily an accurate measure of the overall welfare loss, it does tell us how her misunderstanding of the relationship between $\mathcal{I}_C$ and $\mathcal{I}_S$ contributes to the overall loss. In effect, we learn that if the rest of the consumer's misunderstandings were removed, this is the portion of the overall loss that would remain. If our object is to evaluate an educational

[20]That said, in any given practical application, the evidence may support one of these views more strongly than the other. For example, as reported below, in our setting we find that subjects make simply framed choices much more rapidly than complexly framed choices, which suggests that the former are easier to think about, and therefore potentially more familiar, than the latter. We also find that education regarding pertinent economic concepts changes complexly framed choices, but leaves simply framed ones unaffected. Under the view of the second economist mentioned above, one would expect the opposite pattern.

intervention that aims to improve consumers' understanding of the economic principles governing the relationship between $\mathcal{I}_C$ and $\mathcal{I}_S$, this is a valuable outcome measure.

A final possibility worth considering is that behavior may exhibit additional anomalies not connected to financial education or the complexity of the available alternatives, such as time inconsistency or reference dependence. One can usually construe such patterns as forms of frame dependence (e.g., with time inconsistency, behavior depends on whether choices are made in a forward-looking frame or a contemporaneous one). The Bernheim-Rangel framework allows one to develop normative criteria that recognize multiple frames as welfare-relevant (e.g., both simple contemporaneous and simple forward-looking framing), but this comes at the cost of introducing ambiguity into the evaluation criteria. For an application to the current problem, see Online Appendix A.

3 Experimental Design

We use our notion of financial competence to evaluate a narrow web-based financial education intervention focused on the concept of compound interest.We studied an intervention involving this topic for a number of reasons. First, as noted previously, it is associated with a well-documented bias, and hence is the natural focus of an educational intervention. Second, the design of suitable simply and complexly framed tasks is relatively straightforward. Third, this is a core topic in most financial education courses. Finally, the narrowness of the topic, and the corresponding brevity of treatments in standard investment guides and employer-sponsored financial education programs,[21] make it suitable for an intervention of limited duration. The structure of our experiment also allows us to evaluate the intervention according to both new and conventional outcome measures.

The experiment consisted of three main stages. First, subjects watched one of four educational videos, selected at random. Second, they completed incentivized paired 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 D.

Education intervention We used 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

[21] For example, Skimmyhorn (2015) reports that a financial education program used by the U.S. military covers compound interest 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.

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.[22] 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.[23] 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).

(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.[24] 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. In cases where it was impossible to remove sentences containing rhetorical material, we substituted neutral language.[25] 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

[22] 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."

[23] 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.

[24] These anecdotes do not include any computations, and hence are not helpful for understanding the mechanics of compound interest.

[25] 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?"

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.[26] 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 *Khan Academy*.[27] 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 D.

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 days." Subjects made two sets of choices pertaining to each future reward, one with simple framing, the other with complex framing.[28] 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)).[29]

Table 1 lists the parameters t, r, a, and R used for the paired valuation tasks. We chose these values 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. We chose time horizons of 36 and 72 days to simplify applications of the rule of 72.[30] 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 randomized the order of the valuation tasks at the subject level. Subjects were not told that some of the tasks were

[26]We chose this approach because existing research indicates that financial education videos are generally more effective than written text (Lusardi et al. (2014)).

[27]www.khanacademy.org.

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

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

[30]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.

substantively equivalent, and they typically did *not* perform equivalent simply and complexly framed tasks consecutively.

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, and/or personal advice when making decisions.[31] We take the view that it is appropriate to allow subjects to decide for themselves whether to use such resources, in light of the fact that they always have that option when making real-world decisions. As detailed below, only a quarter of our subjects report exercising that option for answering the incentivized test questionnaire, a fraction that does not vary meaningfully across treatments. Significantly, that pattern mirrors findings concerning real financial decisions (Lusardi and Mitchell (2011)).

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.

Knowledge test and self-reports We also gathered data required 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. (2014), Walstad et al. (2010), Council for Economic Education (CEE) (2006), Collins (2010)). Accordingly, we administered a test consisting of the five questions about compound interest listed in table 2, as well as five ques-

[31] This feature differentiates our study from most of the literature on the effects of financial education (Hastings et al. (2013)). An exception is Levy and Tasoff (ming) who also conduct an internet-based study.

tions about the material covered in the video shown to the control group.[32] As mentioned above, performance on this test was incentivized.[33]

Previous studies have also examined self-reported decision strategies (for instance Heinberg et al. (2014), Lührmann et al. (2012), Carlin et al. (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.[34]

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.

4 Implementation and Preliminary Analysis

Subjects participated in the experiment online rather than in person.[35] We recruited subjects through an online labor market, Amazon Mechanical Turk (AMT). For our purposes, 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 programs in terms of demographic

[32] We randomized the order of the questions at the subject level.

[33] Subjects were informed 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.

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

[35] An advantage of this feature is that it mirrors many real-world financial decisions, which have steadily migrated to online platforms.

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.[36]

We ran eight session with a total of 504 subjects during April and May 2014.[37] 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.[38] 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.[39]

The average length of a session was 62 minutes (s.d. 22 minutes). Attrition was negligible and unrelated to the treatments.[40] 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.[41]

Multiple switching Any subject with coherent preferences will switch his 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, the comparable figure typically falls in the range of 10 to 15%.[42] We drop these subjects from the analysis; the results reported below are based on the 455 subjects who respected monotonicity.

Demographics We provide a detailed analysis of our subjects' demographic characteristics, broken down by treatment group, in the Online Appendix B.1. While our subjects are not highly represen-

[36]We turned to AMT after pilot experiments revealed that the concept of compound interest was already familiar to most Stanford undergraduates. For reviews and evaluations of behavioral research conducted with AMT, see Horton et al. (2011), Mason and Suri (2012), and Peysakhovich et al. (2014).

[37]We ran all of the sessions on weekday mornings.

[38]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.

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

[40]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.

[41]See Mason and Suri (2012).

[42]See, e.g., Holt and Laury (2002).

tative of the US population, neither are they highly unusual. On average, our sample is a 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.[43]

Randomization into treatments was successful. Of the 34 F-tests we performed to test for 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.[44]

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 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, in some cases asking for copies of the videos.[45] Third, in response to the exit survey, the overwhelming majority of subjects reported paying the highest level of attention to the video and to their choices.[46] Fourth, in Online Appendix C.2, we consider the possibilities that inattention may have caused subjects to make noisy choices, or to be unresponsive to varied stimuli. We show that our conclusions are robust with respect to both considerations.

Regression specifications Throughout, we measure treatment effects by estimating the following regression specification:

$$y_{j,k} = \beta_{Control} + \beta_{Full}Full_j + \beta_{Substance}Substance_j + \beta_{Rhetoric}Rhetoric_j + \epsilon_{j,k} \quad (1)$$

[43] Lusardi and Mitchell (2009) report findings based on the American Life Panel, an online survey.

[44] See the Online Appendix B.1 for the results.

[45] In order to control dissemination and exposure of potential future subjects to the videos, we did not provide them in response to these requests.

[46] However, in light of other results reported below, our confidence in the accuracy of self-reports is not high.

Here, j indexes individuals and k indexes decisions; $y_{j,k}$ is an outcome variable, and $Full_j$, $Substance_j$, and $Rhetoric_j$ are treatment dummies. Hence the intercept $\beta_{Control}$ measures the average level of $y_{j,k}$ in the Control condition, and the parameters $\beta_{Full}, \beta_{Substance}$, and $\beta_{Rhetoric}$ measure the effect of the corresponding treatment on that average. We assume $\epsilon_{j,k}$ is orthogonal to the treatment dummies. Whenever a regression includes multiple observations per subject, we cluster standard errors at the subject level.

5 Financial literacy and Self-Reported Decision Strategies

Studies that evaluate financial education interventions frequently focus on financial literacy and/or self-reported behavioral outcomes. (See section 3 for references.) One can draw conclusions about the welfare consequences of financial education from such studies if one is willing to assume that financial education affects behavior only through its effects on the understanding of financial concepts, and that such understanding necessarily promotes better decision making. We begin our analysis by examining effects on a collection of such variables, in each case based on the regression model specified in equation (1), using all pertinent data. All results are highly robust; see the Online Appendix for results based on various alternative specifications.[47]

Column 1 of table 3 shows the effects of the various 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 questions correctly. The Full intervention dramatically increases the average score, by about 1.4 additional correct answers. To put this effect in perspective, the average percentage test score rises from from 39% to 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 effect declines dramatically, to roughly 0.5 (or 10 percentage points).[48] Thus, according to standard measures, the substantive interventions are highly effective at promoting financial literacy.[49]

[47]Robustness analyses in section C.2 of the Online Appendix include the following: adding demographic controls, analyzing behavioral separately for decisions involving each of the two time horizons, analyzing behavior separately for those with high and low initial levels of financial literacy, analyzing behavior separately for subjects with large and small variance in implied rates of time preference in the simply framed choices, and analyzing behavior separately for subjects who react strongly and weakly to changes in experimental stimuli. Finally, Online Appendix section B.3 interacts demographics with treatment effects.

[48]Recall that the Rhetoric-Only treatment includes the simple explanation of compound interest, illustrated through an iterative calculation. Accordingly, it is not surprising to find some effect on measured financial literacy.

[49]Online Appendix C.1 details the effects of the treatments on individual test questions.

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
Level in Control	1.963***	3.284***	0.220***	0.128***	0.0917**	6.404***
	(0.139)	(0.103)	(0.0416)	(0.0396)	(0.0386)	(0.354)
Treatment effects						
Full	1.442***	-1.058***	-0.0126	0.579***	0.172***	1.738***
	(0.197)	(0.146)	(0.0592)	(0.0564)	(0.0550)	(0.504)
Substance-Only	1.271***	-1.339***	0.0611	0.637***	0.260***	1.737***
	(0.189)	(0.140)	(0.0565)	(0.0539)	(0.0525)	(0.482)
Rhetoric-Only	0.492**	-1.079***	0.0655	0.104*	0.0600	0.418
	(0.195)	(0.144)	(0.0584)	(0.0556)	(0.0542)	(0.497)
$P(\beta_{Substance}=\beta_{Rhetoric})$	3.79e-05	0.0618	0.937	0.000	0.000146	0.00606
$P(\beta_{Full}=\beta_{Rhetoric})$	1.73e-06	0.885	0.184	0.000	0.0403	0.00871
$P(\beta_{Substance}=\beta_{Full})$	0.368	0.0467	0.196	0.285	0.0994	0.999
P(joint insignificance)	0.000	0.000	0.400	0.000	3.17e-06	0.000201
Observations	455	455	455	455	455	455
R^2	0.139	0.187	0.007	0.320	0.061	0.043

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Test scores and self reported behavior. The dependent variable in columns 1 - 6 are, respectively, the mean number of exam question answered correctly (1 to 5), the self-reported answer to whether the subjects used external help in the exam, 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, and the self-reported number of complexly framed problems (out of 10) for which the subject explicitly calculated the future reward.

Column 2 of table 3 shows the effects of the various treatments on subjects' test scores for the five test questions pertaining to topics covered in the Control video. Notice that the Control subjects gave slightly more than one additional correct answer on average than the other groups, an improvement of more than 20 percentage points. This finding is notable because it rules out the possibility that differences in test performance between the Control group and the treatment groups are due to effects on general motivation.

Column 3 of table 3 examines the effects of the various educational interventions on the (self-reported) extent to which subjects employ external help. A natural concern is that education may simply displace the use of reference materials or reliance on knowledgable friends. Such displacement could in principle dampen the effects of the interventions on test scores and choices. In fact, the differences across the treatments are small and statistically insignificant. If anything, reliance on external help appears to be slightly higher for the Substance-Only and Rhetoric-Only treatments than for the Control (though the differences are not statistically significant at the 10% level).

The results in Column 4 of table 3 show that subjects reported operationalizing the knowledge they acquired from the substantive interventions. Only 13% of subjects in the Control reported using the rule of 72 when making complexly framed choices. In sharp contrast, the corresponding figure exceeded 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.[50]

As shown in Column 5 of table 3, 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,[51] this finding is not entirely unexpected.

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, the results in Column 6 of table 3 show that the Full and Substance-Only interventions significantly increased the average number of complexly framed decision tasks for which subjects reported 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 was much smaller and statistically insignificant. Thus, the educational interventions did not simply increase (self-reported) reliance on the rule of 72 by migrating subjects from other methods of explicit calculation.

We are of course mindful that changes in self-reported behavior could involve experimenter-demand effects. Indeed, that limitation is an important reason for developing and implementing an objective measure of financial competence. We are much less concerned that similar considerations could account for the measured effects on incentivized test scores, which likely reflect actual knowledge.[52]

Several lessons emerge from this analysis. First, the Full intervention successfully and significantly increased financial literacy. Second, it was successful for the right reason: removing rhetorical material and retaining substance leaves the effect on financial literacy almost unchanged, whereas removing substantive material and retaining rhetoric reduces it dramatically. Third, according to self-reports, the Full intervention successfully motivates subjects to operationalize their newly obtained knowledge in their decisions. Fourth, removing rhetorical material and retaining substance leaves the effect

[50] 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.

[51] 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.

[52] Subjects were plainly motivated to perform well on the incentivized test for their own benefit. Indeed, we received a large number of comments from subjects who complained that they had been tested on material not covered in the intervention video. (Recall that the test covers the substantive material in both the Full video and the Control video, and that each subject views only one video.)

on self-reported operationalization of knowledge almost unchanged, whereas removing substantive material and retaining rhetoric reduces it dramatically. Fifth, according to self-reports, the use of new quantitative tools does not simply crowd out reliance on other resources or other computational methods.

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. However, the results presented in the following sections paint a much different picture.

6 Financial Choices

Much of the literature on financial education is concerned with measuring effects on behavior. (This literature originated with Bernheim and Garrett (2003) and Bernheim et al. (2001); other examples include Duflo and Saez (2003), Bayer et al. (2009), Goda et al. (2012), Cole et al. (2011), Skimmyhorn (2012), Servon and Kaestner (2008), Collins (2010), Lührmann et al. (2014)). Some studies also make casual inferences concerning welfare by asking whether these effects directionally counteract presumed biases and thereby redress deficiencies in decision making. 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, there is a strong presumption that people typically underestimate the power of compound interest (Wagenaar and Sagaria (1975), Eisenstein and Hoch (2007), Stango and Zinman (2009), Almenberg and Gerdes (2012), Levy and Tasoff (ming)), a phenomenon known as *exponential growth bias*. Consequently, following the approach adopted in the literature, one would deem an intervention potentially welfare-improving if it increased the average valuations of the complexly framed rewards.

In this section, we assess treatment effects on valuations by estimating versions of the regression model specified in equation (1), using all pertinent data. All results are highly robust; see the Online Appendix for results based on various alternative specifications.[53]

For the sake of comparability across rewards of different sizes, we normalize the subjects' valuations as follows. For individual j, reward r, time horizon t, and frame $f \in \{simple, complex\}$, the *normalized valuation* is given by

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

[53]Robustness analyses in section C.2 of the Online Appendix include the following: adding demographic controls, analyzing behavioral separately for decisions involving each of the two time horizons, analyzing behavior separately for those with high and low initial levels of financial literacy, analyzing behavior separately for subjects with large and small variance in implied rates of time preference in the simply framed choices, and analyzing behavior separately for subjects who react strongly and weakly to changes in experimental stimuli. Finally, Online Appendix section B.3 interacts demographics with treatment effects.

For simply framed decision problems, we also refer to the normalized valuation as the *rate of time preference*. That interpretation is appropriate if characterization failure is confined to complexly framed problems. Mean rates of time preference are 76.7% and 70.6% for the 36 and 72 day horizons, respectively.[54] The distribution of those rates is spread out between 0% and 100%, with only a few subjects exhibiting rates exceeding 100%.

We start by analyzing treatment effects on normalized valuations in settings with complex framing. Initially restricting our analysis to these settings allows us to illustrate how existing studies of financial education, which typically focus only complexly framed choices, can generate misleading results. Column (1) of table 4 shows that the effect of the Full treatment is large (14.31 percentage points) and statistically significant. Given a presumption that subjects suffer from exponential growth bias, this effect is directionally appropriate. 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.[55] Thus, absent any further analysis, one would be inclined to conclude that the intervention counteracted exponential growth bias, and hence likely made subjects better off. However, as we will see, matters are actually much more complex.

Because we collected data on simply framed choices, we do not need to rely on estimates of exponential growth bias taken from other sources. Instead, we can directly examine the effects of the various treatments on the gaps between valuations in simply and complexly framed settings. First, however, it is important to investigate whether our treatments altered valuations with simple framing. We have hypothesized that subjects evaluate complexly framed options by translating them into simply framed alternatives, and that consequently a misunderstanding of compound interest leads to characterization failure in complexly framed choice problems but not in simply framed ones. If that is the case, we would expect an informative intervention to alter complexly framed choices, but leave simply framed choices unchanged. Column (2) of table 4 shows that, for simply framed valuations, none of the estimated treatment effects are statistically significant at the 5% level, and only the Rhetoric-only treatment is significant at the 10% level. Hence, some of the premises mentioned in our discussion of welfare receive empirical support.

[54] Note that our typical subject discounts future payments rather heavily, that the mean rate of time preference for the longer horizon is lower, and that the relative magnitudes of these rates across horizons are inconsistent with exponential discounting. These three findings are common in studies that elicit time preferences over short horizons (Frederick et al. (2002)).

[55] 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 the formula $FV = I \times (1+i)^{\theta t}$. They estimate this equation for a 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.

VARIABLES	(1) $100 \times \delta^c_{j,r,c}$	(2) $100 \times \delta^s_{j,r,c}$	(3) $100 \times d_{j,r,c}$	(4) $100 \times C_e$	(5) $100 \times C_m$	(6) $100 \times C_e^-$	(7) $100 \times C_e^+$	(8) $100 \times C_m^-$	(9) $100 \times C_m^+$
Level in Control	58.95***	72.26***	-13.31***	11.69***	24.45***	9.241***	2.452***	1.888***	0.557***
	(2.272)	(2.089)	(2.221)	(1.232)	(1.633)	(1.176)	(0.647)	(0.177)	(0.082)
Treatment effects									
Full	14.31***	0.402	13.91***	0.155	-1.584	-4.183***	4.337***	-0.775***	0.616***
	(3.427)	(2.99)	(3.332)	(2.035)	(2.386)	(1.47)	(1.653)	(0.225)	(0.182)
Substance-Only	4.021	0.0182	4.002	-1.461	-2.436	-1.572	0.111	-0.322	0.078
	(3.285)	(2.913)	(2.961)	(1.669)	(2.172)	(1.58)	(0.891)	(0.231)	(0.119)
Rhetoric-Only	18.59***	5.368*	13.22***	-2.546	-4.651**	-5.240***	2.694**	-0.894***	0.429***
	(3.595)	(2.975)	(2.952)	(1.700)	(2.155)	(1.351)	(1.236)	(0.211)	(0.150)
$P(\beta_{Substance} = \beta_{Rhetoric})$	7.96E-05	0.0689	0.000903	0.505	0.270	0.00347	0.0344	0.0025	0.0211
$P(\beta_{Full} = \beta_{Rhetoric})$	0.259	0.0998	0.827	0.177	0.171	0.339	0.375	0.5121	0.3596
$P(\beta_{Substance} = \beta_{Full})$	0.0034	0.897	0.00184	0.413	0.706	0.0585	0.0103	0.0271	0.0036
P(joint insignificance)	1.62E-07	0.202	0.000269	0.390	0.178	0.000251	0.0103	0.0000	0.0007
Observations	4,550	4,550	4,550	4,550	4,550	4,550	4,550	4,550	4,550
Number of subjects	455	455	455	455	455	455	455	455	455
R^2	0.047	0.008	0.033	0.002	0.005	0.019	0.01	0.027	0.018

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Behavioral data. $\delta^f_{j,r,t}$ 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. We set $C_e^- = \big(\min\{d_{j,r,t}, 0\}\big)^2$ and $C_e^+ = \big(\max\{0, d_{j,r,t}\}\big)^2$ to decompose $C_e = C_e^- + C_e^+$ into errors due to under- and overestimation of compound interest, respectively. Similarly, $C_m^- = \big|\min\{d_{j,r,t}, 0\}\big|$ and $C_m^+ = \big|\max\{0, d_{j,r,t}\}\big|$ so that $C_m = C_m^- + C_m^+$. Standard errors clustered by subject.

Now we turn to an analysis of the gaps between matched pairs of simply and complexly framed valuations. We define the *framing distortion* for reward r, time horizon t, and individual j, as:

$$d_{j,r,t} = \delta^c_{j,r,t} - \delta^s_{j,r,t}$$

If an educational intervention eliminates the tendency for subjects to underestimate compound interest, we should find that the average framing distortion is significantly negative for the Control group and zero for the pertinent treatment group.

Column (3) of table 4 presents the effects of the treatments on the mean framing distortion. In the Control condition, subjects' valuations with complex framing are on average 13.31 percentage points lower than with simple framing. The Full treatment increases the average value of $d_{j,r,t}$ by 13.91 percentage points, leaving a gap of only 0.6 percentage points ($s.e. = 2.48$), thereby effectively eliminating the average bias. Significantly, we cannot reject the hypothesis that the mean framing distortion for subjects in the Full treatment is zero at any conventional level of significance.

Before proceeding, it is worth reviewing some of the key results presented so far. We have found that the Full treatment substantially increases measured understanding of compound interest, the frequency with which subjects report operationalizing this understanding when making decisions, and the extent to with which they report making any explicit calculations. Moreover, we find no evidence that the treatment displaces the use of other reliable resources. These findings lead us to conjecture that the Full treatment will not only influence behavior, but will do so for the right reasons. Consistent with that expectation, we find that the Full treatment increases complexly framed valuations, thereby presumably offsetting exponential growth bias. Indeed, the treatment essentially eliminates the average framing distortion.

On the basis of these findings, one would be strongly inclined to conclude that the Full intervention was highly effective and beneficial. Moreover, because the Substance-Only intervention has essentially the same effects on measured financial literacy and self-reported decision strategies, one might well conjecture that it must be equally successful and beneficial, and that rhetoric is an unnecessary distraction. Surprisingly, all of these conclusions are incorrect.

Focusing again on column (3) we see that 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, $p = 0.00184$). In contrast, the estimated effect for the the Rhetoric-Only treatment (13.22 percentage points) is almost identical to that of the Full treatment, and we do not reject equality ($p = 0.827$). Accordingly, despite demonstrable effects on comprehension and subjects' statements concerning their proclivities to operationalize substantive knowledge in their decisions, *the behavioral effects of the Full treatment are traceable almost entirely to its rhetorical components.*

So far, we have focused exclusively on the average framing distortion. Figure 1 shows the cumulative distribution of $d_{j,r,t}$ for each treatment. While framing distortions are clustered near 0, there is substantial variation in $d_{j,r,t}$ within each treatment. Notice that the Substance-Only treatment shifts the CDF for the control group slightly to the right. Both the Full and Rhetoric-Only interventions yield much larger rightward shifts. This pattern is worrisome, inasmuch as an effective educational intervention would shift valuations upward for those with negative values of $d_{j,r,t}$, and *downward* for those with positive values of $d_{j,r,t}$. Thus, one would hope to see the treatment CDFs *crossing* the Control CDF at $d_{j,r,t} = 0$. Instead, the increase in valuation appears to be independent of the initial bias.

In light of these additional findings, there is reason to question whether the Full treatment actually improves the quality of subjects' decisions. By analyzing financial competence, we can address that issue formally and with quantitative precision.

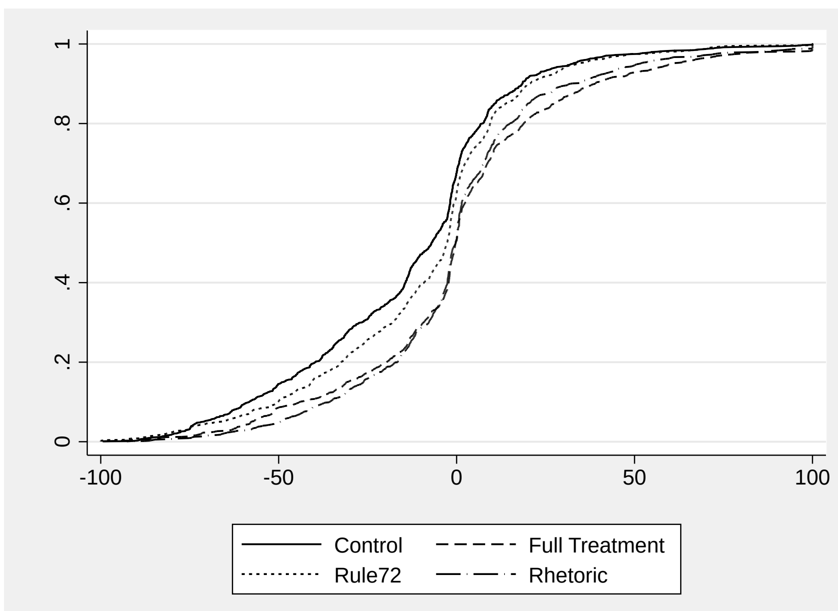

Figure 1: C.D.F. of framing distortion, by treatment. For better visibility, data are truncated at -100 and at 100.

7 Financial Competence

We now turn to the effects of the educational interventions on financial competence. Our general framework equips us with two intuitive and interpretable measures, $C_e = (\delta^s_{j,r,t} - \delta^c_{j,r,t})^2$ and $C_m = |\delta^s_{j,r,t} - \delta^c_{j,r,t}|$. Notice that these measures are positive, and that higher values imply lower competence. If we assume that characterization failure potentially occurs either in simply framed or in complexly framed decision problems, but not in both, then these measures are proportional to the expected and maximal welfare losses in the frame where the failure occurs. Even if characterization failures arise in both types of settings, our measures of competence still have useful welfare interpretations; see section 2.4.

It is natural to assume that poor comprehension of compound interest leads to characterization failure in the complexly framed valuation tasks, but not in the corresponding simply framed tasks. Two separate findings support this assumption. First, subjects take 59 seconds on average to complete one complex valuation task, compared with only 22 seconds for an equivalent simple task ($p < 0.001$). This discrepancy suggests that the complexly framed tasks require additional cognitive effort, presumably because subjects attempt to reduce them to simply framed tasks. Second, as we have already noted, the treatment interventions have significant effects on the complexly framed valuations, but leave the simply framed valuations unchanged on average. We can therefore rule out the possibility that subjects translate simply framed options into complexly framed ones, rather than the other way around. Still, it is worth bearing in mind that the welfare interpretation of our competence measures does not require this assumption.

We assess treatment effects on our two measures of financial competence by estimating versions of the regression model specified in equation (1), using all pertinent data. All results are highly robust; see the Online Appendix for results based on various alternative specifications.[56]. Results appear in columns 4 and 5 of table 4. According to these regressions, the Full treatment has essentially no effect on competence. Indeed, C_e (our measure of average welfare) is actually lower in the Full treatment than in the Control, but the differences are small in magnitude and statistically insignificant. The Substance-Only and Rhetoric-Only treatments are associated with somewhat greater financial competence, but the effects of the Substance-Only treatment are statistically insignificant, as is the effect of the Rhetoric-Only treatment on C_e. Only one treatment effect is statistically signficant: that of the Rhetoric-Only treatment on C_m (the maximal welfare loss).

[56]Robustness analyses in section C.2 of the Online Appendix include the following: adding demographic controls, analyzing behavioral separately for decisions involving each of the two time horizons, analyzing behavior separately for those with high and low initial levels of financial literacy, analyzing behavior separately for subjects with large and small variance in implied rates of time preference in the simply framed choices, and analyzing behavior separately for subjects who react strongly and weakly to changes in experimental stimuli. Finally, Online Appendix section B.3 interacts demographics with treatment effects.

We reach these stunning conclusions despite the fact that the Full treatment significantly enhances financial literacy, 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. The failure of the Full treatment to increase welfare despite these effects is apparently attributable to the two surprising findings noted in the preceding section – that its behavioral effects are in fact driven by rhetoric, and are indiscriminate (in the sense that they are unrelated to the initial framing bias).

We now separate the measure of expected welfare into two components, based on the sign of the framing distortion $d_{j,r,t}$. Specifically, we use $C_e^- = \big(\min\{d_{j,r,t}, 0\}\big)^2$ to capture the part of the welfare loss that is due to *under*estimation of the power of compound interest, and $C_e^+ = \big(\max\{0, d_{j,r,t}\}\big)^2$ to capture the part that is due to *over*estimation. Note that $C_e = C_e^- + C_e^+$. Column (6) of Table 4 reports the effects of the various treatments on C_e^-. As one would expect based on all of the preceding, the Full and Rhetoric-Only treatments significantly reduce welfare losses from underestimating compound interest. (The Substance-Only treatment leaves C_e^- unaffected.) The explanation for the overall null effect of the Full treatment on C_e is readily apparent from column (7): because its behavioral effects are unrelated to the initial framing bias, it *increases* the welfare loss associated with overestimation of compound interest. For those who overestimate, a little learning is a dangerous thing. The two opposing effects roughly offset. For the Rhetoric-Only treatment, the welfare loss associated with overestimation of compound interest also grows, but by a smaller amount. (The Substance-Only treatment also leaves C_e^+ unaffected). Columns (8) and (9) show that similar conclusions follow from an analysis of the maximal welfare loss measures $C_m^- = \big|\min\{d_{j,r,t}, 0\}\big|$ and $C_e^+ = \big|\max\{0, d_{j,r,t}\}\big|$.

These findings are precisely what one would expect in light of Figure 1. An ideal intervention would help each subject recognize the direction and magnitude of his error and adjust accordingly. Yet as noted in the previous section, the Full and Rhetoric-Only treatments appear to raise $\delta^c_{j,r,t}$ indiscriminately, irrespective of whether an individual initially underestimates or overestimates the benefits of compound interest.[57] Visually, both treatments shift the entire distribution of framing distortions to the right. Neither truly resolves exponential growth bias; instead, they appear to introduce countervailing biases. The Rhetoric-Only treatment is more beneficial than the Full treatment because its influence is more highly correlated with the severity of the initial exponential growth bias: it induces a larger rightward shift for small framing distortions, and a smaller rightward shift for large

[57] Levy and Tasoff (ming) also find that a substantial fraction of subjects overestimate compound interest. Song (2015) concludes that the effect of an educational intervention concerning compound interest conducted in rural China on his subjects' retirement savings contributions is largely independent of the extent to which the their saving differed from an optimal benchmark derived from a life-cycle consumption model. In particular, he finds that the education intervention induces some subjects to oversave.

framing distortions (so that the CDFs for the Full and Rhetoric-Only treatments cross). Unfortunately, because the reasons for this fortuitous correlation are unknown, we are unable to draw general conclusions from the apparent superiority of the Rhetoric-Only treatment.

So far, we have proceeded as if each subject's choice mapping is deterministic.[58] Stochastic choice patterns would affect our results through two separate channels. First, were we to substitute the substantively equivalent simply framed valuation task for each complexly framed valuation task in our experiment, so that subjects performed each simply framed task twice, we would likely find that these paired choices would differ in some cases. Applying the method implemented above, we would then measure a positive welfare loss in simply framed choices, even though we would be using (other) simply framed choices as normative benchmarks. As we show in Online Appendix C.3, explicit recognition of this consideration does not alter our main conclusions concerning welfare. Intuitively, noisy choice inflates the overall *level* of the welfare losses measured with our method, but does not materially affect the *relative* magnitudes of the measured welfare losses for the various interventions.

Second, even though stochasticity in choice would not obscure effects on average framing distortions, it could hamper efforts to detect improvements in welfare.[59] We discount this concern because the choices of our typical subject display a high degree of internal consistency.[60] In an abundance of caution, we also address it by performing additional subject-level analyses using a slightly modified welfare measure: for each subject, we compute the average value of $d_{j,r,t}$ over the ten pairs of valuation tasks and square it.[61] Averaging before squaring substantially attenuates the effects of choice stochasticity on measured welfare. Online Appendix C.3 shows that our results remain largely unchanged.[62]

[58] All analyses in this section are based on the assumption that subjects' choices derive from well-defined preferences that satisfy WARP (and are possibly implemented with noise). Online Appendix A presents an analysis of our data based on the Bernheim-Rangel framework which does not require that assumption.

[59] More precisely, if we actually observe $D_{j,r,t} = d_{j,r,t} + \eta_{j,r,t}$, where $\eta_{j,r,t}$ is random noise, we will encounter no bias in measuring the mean of D, or how it changes across treatments. Suppose, however, that the distribution of $d_{j,r,t}$ for the control group is highly concentrated with a mean of $-m < 0$, and that a treatment shifts that distribution upward by a constant amount $c < m$. Under these assumptions, $-(d_{j,r,t})^2$ rises for the vast majority of subjects. However, if the variance of $\eta_{j,r,t}$ is sufficiently large, $D_{j,r,t}$ may be positive for many subjects in the control group, and $-(D_{j,r,t})^2$ will fall for those subjects. With sufficient noise, the latter effect may obscure beneficial effects on welfare.

[60] Online Appendix C.3 shows that subjects who underestimate (overestimate) compound interest in some decisions tend do so in all decisions, and by comparable amounts. Accordingly, the degree of idiosyncratic randomness in individual choice tasks must be limited.

[61] In other words, for each subject j, we calculate $\left(\frac{1}{10}\sum_{r,t} d_{j,r,t}\right)^2$.

[62] One change is that the effect of the Rhetoric-Only treatment on the expected welfare loss is now statistically significant, presumably because subjects who systematically overestimate compound interest are less common than individual choices reflecting overestimation of compound interest.

8 Mechanisms

In this section we shed additional light on the mechanisms through which our interventions affect behavior. First we ask whether they influenced response times for valuation tasks. This question is of interest because slower response times suggest more careful deliberation. Our Substance-only treatment does indeed cause subjects to make complexly framed choices much less quickly. However, when we add rhetoric to the treatment, this effect disappears. We conjecture that rhetoric accelerates these choices by inducing subjects to substitute a less cognitively demand heuristic for rigorous calculations. If that conjecture is correct, it raises the more general possibility that the use of motivational rhetoric may defeat the purpose of substantive instruction. Second, we ask whether the substantive interventions are more effective at improving financial competence for problems to which the rule of 72 is more easily applied. We find no evidence that this is the case. Even for the valuation problems to which the rule is most directly applied (a 1% interest rate over 72 days, and a 2% interest rate over 36 days), the education intervention has no effect on financial competence. Hence, our results are not attributable to a mismatch between the difficulty of the valuation problems and the depth of the material covered in the intervention. Rather, subjects appear to gain knowledge, but fail to apply it when making financial choices. Third, we ask whether our interventions reduce reliance on simple interest calculations. We find that all of them have this effect. Thus, the problem is not one of intellectual stubbornness. Rather, subjects apparently migrate to other similarly inappropriate methods when making their choices.

We start by analyzing the effects of the treatments on response times. Results for complexly and simply framed decision problems appear in columns (1) and (2), respectively, of table 5. As one would expect, we detect no effects for simply framed decisions. Turning to complexly framed decisions, the Substance-only treatment increases the average response time by 19.5 seconds, or roughly 40 percent. 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. Note that adding rhetoric without substance increases response times, possibly because rhetoric alone is motivating but insufficiently instructive.

Next we examine 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-estimated 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.[63] 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 ($p > 0.10$ for all pairwise comparisons).[64] 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.

Lastly, we investigate the effects of our interventions on the extent to which subjects employ simple interest calculations. According to previous research, many people believe investment values grow linearly rather than exponentially (Eisenstein and Hoch (2007), McKenzie and Liersch (2011)).[65] As we show next, *all* of our interventions render this misconception less common.

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:

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

[63] 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 valuations is more likely to be welfare enhancing when the initial bias is greater.

[64] 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.

[65] Likewise, we conducted a pilot study on Amazon Mechanical Turk in which we asked subjects to calculate the future value of four different investments, with given interest rates and time frames. The pilot revealed that the most common modes of calculation are evaluation of simple interest and (correct) evaluation of compound interest. Of course, many subjects fell into neither of these categories.

VARIABLES	(1) $\tau^{c}_{j,r,t}$	(2) $\tau^{s}_{j,r,t}$	(3) $100 \times C_e$	(4) $100 \times C_e$	(5) $100 \times C_e$
Doublings			1	[2 3 4]	2.5
Level in Control	50.81***	22.41***	-7.036***	-13.00***	-12.42***
	(2.675)	(1.073)	(1.435)	(1.409)	(1.909)
Treatment effects					
Full	2.904	-0.768	-0.302	-1.578	5.808
	(4.471)	(1.724)	(1.889)	(2.079)	(3.983)
Substance-Only	19.51***	-0.428	-1.663	-1.914	0.102
	(7.069)	(1.472)	(1.616)	(1.919)	(2.688)
Rhetoric-Only	10.08**	0.941	-2.428	-4.601**	3.502
	(4.599)	(2.197)	(1.589)	(1.829)	(3.310)
$\beta_{Full} - \beta_{Rhetoric}$			1.946	3.023	2.306
$\beta_{Substance} - \beta_{Rhetoric}$			0.765	2.687	-3.400
Observations	4,550	4,550	910	2,730	910
Subjects	455	455	455	455	455
R^2	0.005	0.000	0.006	0.007	0.006

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Response times and problem difficulty. Columns (1) and (2) show the effect of the treatments on mean response times for the complexly and simply framed problems, respectively. Columns (3) - (5) 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. Standard errors clustered by subject.

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.[66] 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 should find $\beta_0^\tau = 1$ and $\beta_1^\tau = 0$, and if all subjects use the simple interest formula, we should find $\beta_0^\tau = 0$ and $\beta_1^\tau = 1$.

We estimate (2) pooling data for all of our subjects, as well as separately for subjects with high and low financial literacy. In each case, we pool data across all valuation tasks.[67] Here we use median

[66]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 $\tilde{\delta}$.

[67]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.

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 0.

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 (roughly 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.[68] 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 formula, they do *not* succeed in fostering the correct calculation of compound interest.

9 Related Literature

Hastings et al. (2013) and Lusardi and Mitchell (2014) provide detailed reviews of the literature on financial literacy and financial education. Here we focus on the portions of that literature that are most closely related to the more novel aspects of our study, as well as some other pertinent areas of inquiry.

Financial education and well-being To our knowledge, Song (2015) is the only existing study that conducts an explicit welfare analysis of a financial education intervention.[69] Farmers in rural China received instruction concerning compound interest, and were then given opportunities to adjust their contributions to a state sponsored retirement savings plan. To evaluate welfare effects, Song employed a life-cycle consumption model parameterized to reflect risk and time preferences elicited from the subjects. He concluded that the intervention improved welfare on average even though its effect on behavior was indiscriminate.[70] While Song's approach allows him to assess the welfare effects

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

[69] Lusardi et al. (2013) use a stochastic life cycle model to theoretically examine the effects of financial education in high school.

[70] Specifically, 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.

of changes in life-cycle consumption plans, it requires him to make a collection of strong assumptions – most notably, that he has the "right" model of life-cycle consumption, and that his preference-elicitation procedure parametrizes it appropriately. In contrast, our approach employs much weaker assumptions.[71]

While avoiding formal welfare analyses, other studies of financial education and the quality of financial decision making have examined various outcome measures that arguably serve as reasonable proxies for non-paternalistic notions of well-being. The general strategy is to examine the effect of financial education on the frequency with which people make choices that are objectively poor irrespective of their preferences (i.e., dominated choices). Studies that have taken this approach include Ernst et al. (2004), Calvet et al. (2007, 2009), Agarwal et al. (2009), Baltussen and Post (2011), and Choi et al. (2011). Choi et al. emphasize the (presumed) dominance relation by using the evocative phrase "dollar bills on the sidewalk."

This approach to evaluating welfare effects has at least three important limitations. First, in naturally occurring settings, dominance is extremely hard to establish, and rationalizations for allegedly poor choices are often possible to imagine. For example, Ernst et al. (2004) point to the use of payday loans by people who have liquidity on credit cards with lower interest rates. But it may be rational for those individuals to avoid depleting all forms of instant liquidity, e.g., to provide for various possible emergencies. Second, even a reduction in dominated choices does not necessarily imply an increase in welfare. For example, indoctrination may help people avoid dominated choices, but only by ignoring their own preferences among undominated alternatives. Indeed, we have seen in the current study that suppressing a particular type of dysfunctional behavior (here, the use of simple interest calculations) need not improve welfare, because new choice patterns may be equally problematic. Third and finally, one cannot translate effects on the frequency of dominated choices into standard welfare measures such as compensating or equivalent variations, which are needed for cost-benefit analyses of potential interventions.

Experimental evaluation of narrow financial education interventions Our experiment examines a narrow educational intervention that focuses on a particular skill, and evaluates it based on test questions and decision tasks that are directly connected to its substantive content. In contrast, much of the existing literature on financial education examines broad, highly composite, and often heterogeneous programs (such as high school classes or workplace seminars), as well as behaviors that the curricula may not explicitly address (such as rates of saving).[72] We suspect that these considerations may account for the literature's mixed findings concerning the effects of financial education.

[71] Of course, unlike Song's approach, it does not provide measures of life-cycle welfare.

[72] See for instance Bernheim et al. (2001), Bernheim and Garrett (2003), Mandell (2009), and Cole and Shastry (2012).

The current study adds to a burgeoning literature that demonstrates the importance of investigating the narrow constituent parts of such interventions using experimental methods. Goda et al. (2012) conduct a large field experiment in which employees of the University of Minnesota are provided with information about voluntary retirement savings plans. In one treatment, different savings levels are projected into assets at retirement. Another treatment adds projections of retirement income.[73] The authors find significant increases in contributions when income projections are provided, but none when projections are limited to assets at retirement. Like the current paper, their research demonstrates that seemingly minor differences in the provision of information sometimes have large behavioral effects, and that identifying the drivers of behavioral change is critical for the design of effective interventions. In a similar vein, Drexler et al. (2014) provided a group of micro-entrepreneurs in the Dominican Republic with basic training in accounting, and compared the effectiveness of this intervention with training that emphasizes rules of thumb and heuristics. For less highly skilled subjects, the rules-of-thumb training led to significantly greater improvements in firms' financial practices, objective reporting of quality, and revenues. In an unincentivized experiment, Carlin et al. (2014) found that subjects' propensity to choose the best credit card from a list of options is significantly enhanced when explicit information about the location of the pertinent information is added to an educational intervention. Finally, Heinberg et al. (2014) administered an educational intervention concerning five basic financial concepts, in some cases through written narratives and in others through videos. They found that, of the two approaches, videos more effectively improve motivation and perceived self-efficacy regarding financial decision making. Likewise, Lusardi et al. (2014) find that videos are more effective at improving financial literacy.

Imperfect decision making As mentioned in Section 2, the current study is an application of behavioral welfare economics, and therefore can be read as a contribution to that burgeoning literature. It is also related to other research that explores aspects of imperfect decision making.

Our study is also related to a handful of papers that investigate the effects of complex framing on decision making. Hastings and Tejeda-Ashton (2008) investigate hypothetical choices among investment funds by financially illiterate Mexican workers, and show that subjects are more inclined to select funds with lower fees when those fees are presented as pesos rather than annual percentage rates. In a field experiment, Bertrand and Morse (2011) find that providing borrowers with information that reinforces the adding-up of dollar fees incurred when rolling over loans reduces the take-up of future payday loans by 11%. Kalaci and Serra-Garcia (ming) conduct an experiment in which subjects have to choose from a set of options that entail both costs and benefits. They find that complex presentation of the costs increases subjects' propensity to choose the highest gross-benefit

[73] In both treatments, projections are customized to the recipient's financial situation.

option (which differs from the highest net-benefit option), whereas complex presentation of the benefits induces more random choice (rather than increasing subjects' propensity to choose the lowest-cost option). Abeler and Jaeger (2014) study subjects' effort choices in a piece rate task involving taxes and subsidies that are framed either simply or complexly. They find that complex framing reduces the magnitude of responses to changes in tax rates, compared with simple framing. In contrast to our study, they evaluate complexly framed choices in relation to a theoretical benchmark, rather than in relation to each subject's own simply framed choices.[74]

10 Conclusion

In this paper, we have introduced the notion of *financial competence*, and used it to analyze the effects of a financial education intervention concerning compound interest.

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 a consumer's decisions across equivalent complexly framed and simply framed valuation tasks that lead to the same intermediate outcomes. As a method of evaluating the quality of financial decision making, this approach offers a number of significant advantages over conventional metrics: it is non-paternalistic; it yields a quantitative measure of the quality of financial decision making which, under relatively modest assumptions, are formally interpretable in terms of consumer welfare; it imposes modest information requirements; and it explicitly and flexibly accounts for population heterogeneity, which is a key consideration when evaluating the effects of financial education, due to differences in initial knowledge and misinformation.

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.

An examination of two additional interventions (one without rhetoric, one with limited substance) reveals that while the effects on measured financial literacy and self-reported choice strategies are attributable to the substantive components of the intervention, changes in behavior are almost *en-*

[74]Each of their subjects makes either complexly or simply framed choices, not both. Hence, their design precludes the use of a subject's simply framed choices as normative benchmarks for their complexly framed choices. They do, however, compare subjects' earnings in the complexly framed treatment not only to the theoretical benchmark, but also to other subjects' earnings in the simply framed treatment, and they argue that differences in effort costs across subjects are unlikely to affect behavior in their setting.

tirely attributable to the rhetorical components. As a result, despite the intervention's success in discouraging subjects from employing simple interest calculations, it does not induce them to evaluate compound interest correctly when making decisions (even though tests reveal that their ability to compute compound interest improves dramatically). Thus, while the intervention appears highly successful according to conventional measures, it is not actually beneficial.

In generalizing from these results, one must of course exercise caution, especially since the interventions we study are limited. The most important lesson to be drawn from this analysis is not that a particular intervention had certain effects, but rather that it is possible for financial education to be highly successful according to conventional outcome measures while failing to improve the quality of financial decision making. Thus, we provide a decidedly negative answer to an important open question identified in the literature review by Hastings et al. (2013): "whether test-based measures provide an accurate measure of actual financial capability." While we remain convinced that financial literacy is important, it does not by itself guarantee financial competence.

Our main findings pose serious challenges for public policy regarding financial education. At the strategic level, we can imagine three broad alternative approaches. The first is to devise educational methods that more effectively lead people to put pertinent knowledge into practice, and to do so correctly, when they make decisions. Implementing this strategy will require extensive research into the effects of alternative pedagogical approaches not only on financial literacy, but also on financial competence. Given that brevity appears to be a design constraint for employer-based financial education, it is important to determine whether efficacy and brevity are compatible. A second approach is to deploy educational programs targeted at populations known to manifest particular biases in order to create countervailing biases (in effect accomplishing the right objective for the wrong reason). For example, in the current study, we have found that the most beneficial intervention is the one with the least substance and the most emphasis on rhetoric. 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. We are skeptical about the practical value of this "debiasing" approach, because it seems likely that any success in balancing countervailing biases will be context-specific, and consequently not necessarily indicative of how any particular individual would make a broad range of real-world decisions involving the pertinent financial concepts. A third approach is to develop better tools to assist with real-world decision making. Using our terminology, the object would be to turn naturally occurring complexly framed decision problems into simply framed problems. Similarly, legislation could require suppliers of financial products to characterize them in simply framed terms.[75] In principle this is a promising approach, but its effective deployment 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 educational interventions in the workplace, as well as in high schools. Research on pedagogical design will, however, at least initially require extensive study of more narrowly focused interventions in the laboratory. Indeed, we have emphasized that 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 think the call by Hastings et al. (2013) for studies of "large scale interventions" may be premature. The effective design of large-scale interventions likely requires a much more comprehensive micro-level understanding of financial education than we currently possess. An initial focus on narrowly focused, 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.

[75] As the current research demonstrates, the disclosure of the annualized percentage interest rate, as required by the Truth in Lending Act, may not achieve that objective.

References

Abeler, J. and Jaeger, S. (2014). Complex tax incentives. *American Economic Journal: Economic Policy*.

Afriat, S. N. (1972). Efficiency estimation of production functions. *International Economic Review*, 13(3):568–98.

Agarwal, S., Driscoll, J. C., Gabaix, X., and Laibson, D. (2009). The age of reason: Financial decisions over the life cycle and implications for regulation. *Brookings Papers on Economic Activity*, Fall:51–101.

Almenberg, J. and Gerdes, C. (2012). Exponential growth bias and financial literacy. *Applied Economics Letters*, 19(17).

Andersen, S., Harrison, G. W., Lau, M. I., and Rutstrom, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9:383–405.

Arrow, K. J. (1959). Rational choice functions and orderings. *Economica*, 26(102):121–127.

Baltussen, G. and Post, G. T. (2011). Irrational diversification: An examination of individual portfolio choice. *Journal of Financial and Quantitative Analysis*, 5:1463 – 1491.

Barnett, S. M. and Ceci, S. J. (2002). When and where do we apply what we learn? a taxonomy for far transfer. *Psychological Bulletin*, 128(4):612.

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

Bernheim, B. D. (2009). Behavioral welfare economics. *Journal of the European Economic Association*, 7(2-3):267–319.

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

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

Bernheim, B. D. and Rangel, A. (2004). Addiction and cue-triggered decision processes. *American Economic Review*, 94(5):1558–90.

Bernheim, B. D. and Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics*, 124(1):51–104.

Bertrand, M. and Morse, A. (2011). Information disclosure, cognitive biases, and payday borrowing. *The Journal of Finance*, LXVI(6):1865–93.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *Quarterly Journal of Economics*, pages 1243–85.

Calvet, L. E., Campbell, J. Y., and Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy*, 115(5):707–47.

Calvet, L. E., Campbell, J. Y., and Sodini, P. (2009). Measuring the financial sophistication of households. *American Economic Review: Papers & Proceedings*, 99(2):393–398.

Carlin, B. I., Jiang, L., and Spiller, S. A. (2014). Learning millennial-style. *Working Paper, Anderson School of Business, UCLA.*

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

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

Choi, S., Kariv, S., Mueller, W., and Silverman, D. (2014). Who is (more) rational? *American Economic Review*, 104(6):1518–1550.

Cole, S., Sampson, T., and Zia, B. (2011). Prices or Knowledge? What Drives Demand for Financial Services in Emerging Markets? *The Journal of Finance*, 66(6):1933–1967.

Cole, S. and Shastry, G. K. (2012). Is high school the right time to teach self-control? The effect of education on financial behavior. *Unpublished Manuscript, Harvard University.*

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

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

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

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

Echenique, F., Lee, S., and Shum, M. (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6):1201–1223.

Eisenstein, E. M. and Hoch, S. J. (2007). Intuitive compounding: Framing, temporal perspective, and expertise. *Unpublished Manuscript*.

Ernst, K., Farris, J., and King, U. (2004). Quantifying the economic cost of predatory payday lending. Technical report, Center for Responsible Lending.

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

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401.

Goda, G. S., Manchester, C. F., and Sojourner, A. (2012). What will my account really be worth? an experiment on exponential growth bias and retirement saving. *NBER working paper*, 17927.

Grether, D. M. and Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69(4):623–38.

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

Hastings, J. S. and Tejeda-Ashton, L. (2008). Financial literacy, information, and demand elasticity: Survey and experimental evidence from mexico. *NBER Working Paper*, 14538.

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

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644 – 1655.

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14:399–425.

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

Kalaci, K. and Serra-Garcia, M. (Forthcoming). Complexity and biases: An experimental study. *Experimental Economics*.

Kline, P. (1999). *Handbook of Psychological Testing*. Routledge, London and New York, 2 edition.

Levy, M. R. and Tasoff, J. (Forthcoming). Exponential growth bias and life cycle consumption. *Journal of the European Economics Association*.

Lührmann, M., Serra-Garcia, M., and Winter, J. (2012). Teaching teenagers in finance: Does it work? *Munich Discussion Paper*, 24.

Lührmann, M., Serra-Garcia, M., and Winter, J. (2014). The impact of financial education on adolescents' intertemporal choices. *Unpublished Manuscript, University of Munich*.

Lusardi, A., Michaud, P.-C., and Mitchell, O. S. (2013). Optimal financial knowledge and wealth inequality. Technical Report 18669, National Bureau of Economic Research Working Paper.

Lusardi, A. and Mitchell, O. (2009). How ordinary consumers make complex economic decisions: Financial literacy and retirement readiness. *NBER Working Paper*, 15350.

Lusardi, A. and Mitchell, O. (2011). Financial literacy and planning: Implications for retirement well-being. In Lusardi, A. and Mitchell, O., editors, *Financial Literacy. Implications for Retirement Security and the Financial Marketplace*, pages 17–39. Oxford University Press.

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

Lusardi, A., Samek, A. S., Kapteyn, A., Glinert, L., Hung, A., and Heinberg, A. (2014). Visual tools and narratives: New ways to improve financial literacy. *NBER Working Paper*, 20229.

Malkiel, B. G. and Ellis, C. D. (2013). *The Elements of Investing. Easy Lessons for Every Investor.* Wiley, New Jersey.

Mandell, L. (2009). 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*, Washington, D.C.

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

Mason, W. and Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavioral Research*, 44:1–23.

McKenzie, C. R. M. and Liersch, M. J. (2011). Misunderstanding savings growth: Implications for retirement savings behavior. *Journal of Marketing Research*.

Peysakhovich, A., Nowak, M. A., and Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature Communications*, 5.

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

Skimmyhorn, W. L. (2012). *Essays in behavioral household finance*. PhD thesis, Harvard Kennedy School, Cambridge, MA.

Skimmyhorn, W. L. (2015). Assessing financial edcuation: Promising evidence from boot camp. *USMA Working Paper*.

Song, C. (2015). Financial illiteracy and pension contributions: A field experiment on compound interest in china. *Unpublished Manuscript*.

Stango, V. and Zinman, J. (2009). Exponential growth bias and household finance. *Journal of Finance*, 64(6).

Tversky, A., Slovic, P., and Kahneman, D. (1990). The causes of preference reversal. *American Economic Review*, 80(1):204–17.

van Rooij, M., Lusardi, A., and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, 101:449–472.

Wagenaar, W. M. and Sagaria, S. D. (1975). Misperception of exponential growth. *Perception and Psychology*, 18(6):416–422.

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