Gutenberg School of Management and Economics

& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

Pop-ups Pay Off: Simulating App-Based Trading to Boost Financial Competence

Gabriele Iannotta, Katharina Hartinger, Tommaso Agasisti

6th May 2026

Discussion paper number 2603

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
https://wiwi.uni-mainz.de/

Contact details

Gabriele Iannotta
Department of Management, Economics and Industrial Engineering
Politecnico di Milano
Via Lambruschini 4
20156 Milano
Italy
gabriele.iannotta@polimi.it

Katharina Hartinger
Chair of Public and Behavioral Economics
Johannes Gutenberg University
Jakob-Welder-Weg 4
55128 Mainz
Germany
kharting@uni-mainz.de

Tommaso Agasisti
Department of Management, Economics and Industrial Engineering
Politecnico di Milano
Via Lambruschini 4
20156 Milano
Italy
tommaso.agasisti@polimi.it

Pop-ups Pay Off: Simulating App-Based Trading to Boost Financial Competence *

Gabriele Iannotta[a], Katharina Hartinger[b], and Tommaso Agasisti[a]

[a]Politecnico di Milano, Department of Management, Economics and Industrial Engineering
[b]Johannes Gutenberg University Mainz

May 06, 2026

Abstract

The advent of commission-free trading apps has drawn millions of young, financially inexperienced users into capital markets, raising concerns about their preparedness to navigate behavioral pitfalls embedded in platform design. We evaluate two short and scalable simulation-based financial education interventions in a three-arm randomized experiment with 704 undergraduate students at an Italian university (488 completers). In both treatments, participants trade fictitious assets in an incentivized 20-round game that simulates a trading-app environment, accompanied by introductory educational content on core investment concepts. The augmented treatment additionally embeds short in-game pop-ups addressing behavioral pitfalls relevant to app-based trading, including diversification, overtrading, the disposition effect, availability bias, and herd behavior. Measured two weeks after the intervention, both treatments significantly increase financial knowledge relative to a no-intervention control group, with effect sizes of approximately 0.25-0.30 SD for the baseline simulation and about 0.5 SD for the pop-up-augmented version. Both treatments also improve portfolio efficiency captured by a design-based Sharpe ratio computed from declared allocations, while the augmented treatment additionally increases realized in-game portfolio efficiency and revealed risk-taking during the incentivized simulation. By contrast, stated risk attitudes remain unchanged, indicating that the intervention improves how financial knowledge is translated into portfolio decisions rather than altering underlying risk preferences.

Keywords: Financial education; Financial literacy; Trading apps; Portfolio efficiency; Learning-by-doing; Behavioral nudges; Randomized controlled trial.
JEL classification: G53; G11; G41; C93; I21.

---

*We thank Ilja Cornelisz, Daniel Schunk, David Streich, Chris van Klaveren, Simon Wiederhold, seminar audiences at JGU Mainz and the CLE Seminar, and conference participants at the 13th BEEN, the 50th and 51st AEFP Conferences, AQMAPPS II and III, the Prague Conference on Behavioral Sciences, the IX INVALSI Seminar, and the 11th IWEEHPS for helpful comments and suggestions. Ethical approval was obtained by the IRB of the Politecnico di Milano. All errors are ours.

1 Introduction

Improving financial behaviors is widely considered the ultimate goal of financial education (FE) interventions. The persistence of bad financial habits highlights the need to support young adults in making better financial decisions (West and Cull, 2020). This challenge is especially salient in light of rapid changes in the financial sector, where trading platforms have gained substantial traction in recent years. The scale of this shift is substantial: the rise of commission-free, mobile-first brokerages has drawn millions of first-time investors into financial markets, with growth accelerating sharply during and after the COVID-19 pandemic (Barber et al., 2022). Regulatory authorities have flagged the particular vulnerability of young, inexperienced users to the behavioral risks embedded in these platforms (Brescia Morra et al., 2025). Young adults are disproportionately represented among these new participants, many of whom begin investing with limited financial knowledge and little prior experience with capital markets (Blake et al., 2022; Lusardi and Mitchell, 2014). Yet, the same platforms that broaden access to investing also introduce new challenges for financial decision-making. They often encourage frequent trading and incorporate gamified elements—as documented for platforms such as Robinhood and Trade Republic—including awarding badges and organizing challenges, thereby blurring the line between trading and gambling. As Chapkovski et al. (2026) and Chapkovski et al. (2025) demonstrate, gamified interfaces indeed increase engagement, but also noisy trading behaviors among retail traders.

This paper argues that FE holds considerable promise for mitigating the dangers posed by these platforms, provided it adapts productively to these new developments in the financial sector and equips users with the competence needed to navigate them effectively. The accessibility of app-based trading has outpaced the financial preparedness of its users: young adults can now enter capital markets in seconds, yet frequently lack the knowledge required to evaluate risk, interpret performance signals, or recognize behavioral traps embedded in platform design. In this context, financial education is not merely about transmitting abstract principles; it must bridge the gap between the technological ease of market access and the decision-making competence that informed participation demands. Specifically, we evaluate two fast-paced, light-touch, and highly scalable learning-by-doing education concepts in an experiment with undergraduate students at an Italian university. In total, 704 students were randomized into treatment and control groups after completing a pre-questionnaire, and 488 completed the full study protocol online and constitute the completion sample. Students in the treatment arms participated either in a simulation game resembling a trading app or in the same game augmented by short educational messages focusing on behavioral pitfalls and decision-making frictions that are salient in app-based trading. We find that the baseline learning-by-doing simulation improves financial knowledge and behaviors, but the augmented simulation with behaviorally informed pop-up prompts produces significantly larger gains in financial knowledge and in realized in-game portfolio efficiency, while yielding statistically comparable improvements in declared portfolio efficiency. Specifically, participating in the simulation game raises financial knowledge test scores by about 7-8% relative to the control mean and improves the Sharpe ratio for a hypothetical portfolio allocation task by 27%. These effects are measured two weeks after the intervention. In contrast, receiving short (non-personalized) educational pop-ups during the simulation game raises knowledge by about 15-16% and the Sharpe ratio by 38% compared to the control group. When comparing incentivized behavior *within the simulation game* between the two treatment groups, a similar picture emerges: the in-game Sharpe ratio improves by 51% for the pop-up group relative to T1—highlighting a strong complementarity between learning-by-doing and appealingly presented in-game education content, even when this content is not designed to provide a solution to the game but rather to illustrate the general behavioral challenges associated with trading apps.

These main results hold when we add extensive individual controls as well as include Lee (2009) bounds to account for attrition. We do not find systematic evidence of treatment-effect heterogeneity within the student sample. While we—as expected—do not find an effect of the interventions on stated risk attitudes, we find that the pop-up-augmented treatment positively affects revealed risk-taking: participants in T2 allocate a significantly higher share of their budget to risky assets than participants in T1.

The simulation-based, pop-up-augmented intervention builds on two strands of the financial literacy literature: evidence on the limited translation of financial knowledge into behavior, and evidence on the potential of gamified learning-by-doing approaches in this context. Although traditional research has often equated financial knowledge with overall financial competence, Ambuehl et al. (2014) caution that even when individuals acquire financial principles, they may not implement them—or may deploy them incorrectly. This concern is especially pronounced in the context of trading apps, with their distinctive psychological and behavioral dynamics (Chaudhry and Kulkarni, 2021), and with financial novices as an important target group of such platforms (Chapkovski et al., 2026). While quantitative skills form a fundamental component of financial capability, recent studies (Carpena and Zia, 2020; Horn et al., 2023) indicate that these skills play only a limited role in real-world financial decision-making. Instead, the key challenge lies in bridging the gap between financial knowledge and behaviors. Carpena et al. (2019) emphasize the importance of understanding the psychological and contextual barriers that prevent individuals from translating knowledge into effective action, while Yoong (2011) advocates linking FE directly to concrete financial behaviors. This perspective is further reinforced by Amagir et al. (2018), who argue for integrating behavioral economics into FE programs to more effectively target the cognitive biases and decision-making pitfalls inherent in complex financial environments.

A key implication of this evidence is that the effectiveness of FE depends not only on "what" is taught but on "how" it is delivered. If the central challenge lies in translating knowledge into behavior, then the pedagogical approach—and, in particular, the degree to which learners are actively engaged in realistic decision-making contexts—becomes a first-order design parameter. A growing body of interventions grounded in active learning principles, particularly experiential learning, reflects this insight. As noted by Kaiser and Menkhoff (2020), when educational programs are made more entertaining or personalized, their impact on financial behaviors is significantly enhanced. Issues like digital versus traditional education formats are often discussed when looking at FE for younger cohorts, such as Gen Z (Sconti, 2022). Lobão (2020) further contends that experiential learning activities, such as virtual market participation and trading simulations, allow participants to directly experience specific situations—both in content and in context. In addition, experimental games help overcome limitations of traditional survey methods, which often capture only self-reported data. By incorporating appropriate incentives, such games encourage participants to engage fully and exert the necessary effort to solve assigned tasks (Muñoz-Murillo et al., 2020). In a recent pilot RCT with German undergraduate students, Oberrauch and Kaiser (2024) find that a brief video-based financial education program produces large and persistent knowledge gains (about 0.5 SD), whereas subsidized access to a robo-advised ETF platform—a pure learning-by-doing channel—yields zero effects on financial knowledge, and the two interventions show no complementarity. Our findings nuance this conclusion: the simulation alone (T1) does produce measurable gains, but the largest effects emerge when experiential learning is augmented by short educational prompts that explicitly address behavioral pitfalls.

We contribute to these strands of literature in two main ways: (i) by providing a rigorous experimental evaluation of scalable FE interventions that target core aspects of financial literacy in the modern trading app context, and (ii) by offering new insights into how FE can translate financial knowledge into more effective portfolio choices in realistic, payoff-relevant investment environments by integrating a gamified learning-by-doing approach with educational messages that explicitly address behavioral biases and challenges associated with relatively new financial technologies.

Financial education encompasses a wide range of topics—from budgeting and saving to debt management, insurance, and retirement planning. We focus on asset allocation because it lies at the core of the decision problem that trading apps present to their users: how to distribute wealth across risky and risk-free instruments under uncertainty. A natural entry point for this analysis is the widespread difficulty of translating basic portfolio principles into effective allocations: for example, the Linciano et al. (2019) report reveals a widespread lack of diversification knowledge among Italian households—a finding corroborated by studies from Mouna and Jarboui (2015), Abreu and Mendes (2010), and Guiso and Jappelli (2009), which document a direct link between low financial literacy and under-diversified portfolios. At the same time, and crucially for the trading-app context, the relevant competence is not simply "diversifying more" in the sense of maximizing the number of assets held. In digital choice architectures, superficially diversified portfolios can remain inefficient if positions are redundant in terms of co-movement, while efficient risk management can sometimes be achieved with fewer positions when correlation and downside risk are understood. Consistent with this, our behavioral outcomes are framed around portfolio efficiency—that is, the extent to which a given allocation achieves the highest expected return for the level of risk taken, as captured by the ratio of portfolio expected return to portfolio volatility—rather than around the count of holdings.

Beyond portfolio construction, the study investigates the role of risk aversion—a central but empirically ambiguous factor in financial decision-making. Existing evidence provides mixed conclusions: some studies suggest that greater financial literacy may heighten risk aversion (Amari et al., 2020; Dinç Aydemir and Aren, 2017; Litterscheidt and Streich, 2020), while others indicate that it may instead reduce it (Aren and Zengin, 2016; Bajo et al., 2015; Bayar et al., 2020; Hermansson and Jonsson, 2021). A key reason for this ambiguity lies in how risk aversion is measured. Much of the literature relies on self-reported indicators, which capture individuals' stated attitudes toward risk, but may not fully reflect how they behave when facing actual investment choices. By explicitly combining validated self-reported measures of risk attitudes with observed allocation choices made in an incentivized trading simulation, this study goes beyond standard survey-based approaches. It provides direct evidence on whether FE affects not only stated risk preferences, but also actual openness to holding risky assets in an incentivized repeated investment decision.

Furthermore, our research addresses asset allocation more broadly. Although the share of retail investors in European assets under management has grown steadily in recent years—rising from 26% in 2020 to nearly 32% by end 2024, driven partly by the emergence of user-friendly digital platforms targeting young investors—European households continue to hold a disproportionate share of their savings in bank deposits, forgoing substantial long-term returns (European Fund and Asset Management Association (EFAMA), 2025). In a series of interviews, Duraj et al. (2025) highlight the role of perceived inaccessibility in keeping retail investors out of financial markets—an obstacle that trading apps aim at overcoming. Consequently, it is of first-order importance to understand how FE can promote both enhanced financial knowledge and improved decision-making competence within these new, digitally mediated financial environments—with the aim of preventing both excessive risk-taking and the opposite pattern of excessive reliance on guaranteed savings vehicles. Overall, by integrating theoretical insights from behavioral finance with an experimental design that captures both self-reported and observed measures, our study advances the discussion on how FE can be most effectively deployed to enhance financial decision-making in a rapidly evolving digital landscape.

The remainder of the paper is organized as follows: Section 2 presents the educational framework underlying the interventions. Section 3 describes methodology and research questions and is followed by Section 4, which details the simulation game. Section 5 discusses the empirical approach and main results, while Section 6 presents the attrition analysis. Section 7 concludes.

2 Education concept and intervention design

We model the trading app simulation as an opportunity for participants to invest in their human capital—specifically, their FE. When applying the simple human capital framework (Becker, 1962; Lusardi et al., 2017) to a learning-by-doing simulation game, the participant weighs the costs and future benefits of training. Specifically, conditional on participating in the simulation, the decision made by the participant is whether to use the task as a learning opportunity and invest an appropriate amount of effort. The benefits of investing effort are twofold: an immediate monetary benefit arises from success in the simulation game itself, which determines the participant's payoff. In addition, as in a typical education decision, there are long-run returns to education. Unlike in the classical case, these returns do not occur in the labor market but in capital-market decisions when participants use trading apps or similar digital investment interfaces. Since the investment game is a stochastic game, the decision to invest effort does not follow mechanically from the set-up, but can plausibly be seen as a deliberate (albeit potentially subconscious) choice made by the participant.

Zooming in on the design of the simulation game illuminates the outcomes on which we expect an effect as well as the underlying mechanisms. The core idea behind the trading app simulation game is a learning-by-doing approach. Rather than teaching financial literacy (only) in the context of a regular class, the simulation game allows users to actively participate in trading, observe consequences of their investment choices over multiple rounds, and gain a deeper understanding of the stochastic nature of trading. Crucially, the simulation bridges the gap between completely risk-free trading in a FE exercise—since the simulation game will affect the final payoff—and real-stakes trading on a real-world trading platform. The close link between in-simulation behavior and app-like investment decisions allows participants to translate new knowledge directly into portfolio choices within the task. Thus, we expect the trading simulation to affect both financial knowledge and financial behaviors.

This effect should be further augmented by the pop-up treatment in the respective group, as pop-ups reinforce the pedagogical nature of the simulation task. The pop-ups are designed as brief educational prompts that increase the salience of common behavioral mechanisms relevant to investment decision-making within an app-based trading environment at moments when participants are actively engaged with investment decisions[1]. Some pop-ups convey general investment principles, such as the importance of diversification, while others address well-documented behavioral biases that affect trading behavior. Importantly, two of the five messages focus explicitly on phenomena that are particularly salient in trading app contexts—overtrading and herd behavior—arising from platform features that encourage frequent transactions and social comparison among investors. Due to their general and practical nature, the pop-ups have very limited direct relevance in the context of the simulation game: they do not disclose any additional information about the game's payoff structure (e.g., probabilities, correlations, or optimal allocations), nor do they provide round-specific recommendations. In this sense, they are not intended to function as performance guidance or "solutions" to the task. Their brevity makes them invitations to explore specific aspects of financial skill and trading app use in greater depth, rather than comprehensive lessons. This design aligns with the concept of "behavioural-based financial education" introduced by Pitthan and De Witte (2025), who show that financial education materials explicitly targeting cognitive biases can improve financial literacy not only directly but also indirectly, through increased awareness of the biases themselves. In their RCT with secondary school students, the indirect channel accounts for a significant share of the total treatment effect—but only when the course materials explicitly address the behavioral mechanism, not when they follow a traditional curriculum. Our pop-up design follows this logic by drawing attention to the specific pitfalls described above rather than providing generic financial instruction. Pop-ups are well suited for capturing attention: their element of surprise parallels marketing tools designed to cut through information overload, and recent economic research has emphasized how attention constraints shape decision-making. We leverage this mechanism to convey messages about non-technical aspects of using trading apps in an empathetic way. A further pop-up design feature is central to our education concept. The prompts are exogenously timed: they appear in pre-determined rounds irrespective of prior choices or performance, provided that the participant reaches the relevant round. In practice, variation in exposure due to early termination is limited: the

[1]Appendix C reports the full wording of the pop-up messages in the order of their first appearance and reproduces an original in-game screenshot of the diversification pop-up.

median number of rounds completed is 20 in both treatment arms, with means of 18.7 (T1) and 19.2 (T2) (see Table 14 in Appendix B). Thus, they do not provide dynamic feedback in any way, distinguishing our intervention from the literature on personalized real-time feedback or adaptive learning (Aucejo and Wong, 2025; Brade et al., 2022; Muralidharan et al., 2019; Van Klaveren et al., 2017). In contrast to, say, a personal one-on-one meeting with a financial advisor, the pop-up-augmented simulation is a cost-efficient and easily scalable intervention to increase knowledge and improve financial decision-making. In other words, the pop-ups provide a lightweight educational scaffold at the point of decision—a role that, in real-world trading, is typically left unfilled for retail investors who operate without professional advice.

3 Methodology and research questions

To rigorously evaluate the effectiveness of our FE interventions, we conducted an online randomized experiment with undergraduate students at Politecnico di Milano. Ethical approval was obtained through Politecnico's IRB. The experiment was not pre-registered. To ensure transparency, we provide a dedicated appendix (Appendix A) documenting the experimental design, outcome construction, and the full set of specifications reported in the paper. Recruitment began with an email invitation, accompanied by an informational flyer, sent to all eligible students. Participants were directed to a dedicated online platform, where they first completed a pre-questionnaire from November 22 to 26, 2023. This questionnaire established baseline levels of financial knowledge, attitudes, and self-reported behaviors. After completing and submitting the pre-questionnaire, students were randomly assigned to one of three experimental arms (see Figure 1). Randomization was implemented only after the baseline questionnaire had been submitted, so participants completed all pre-treatment measures without knowing their assigned group. This design ensures baseline comparability across arms in expectation, allowing differences in outcomes to be attributed to the intervention rather than to pre-existing individual characteristics.

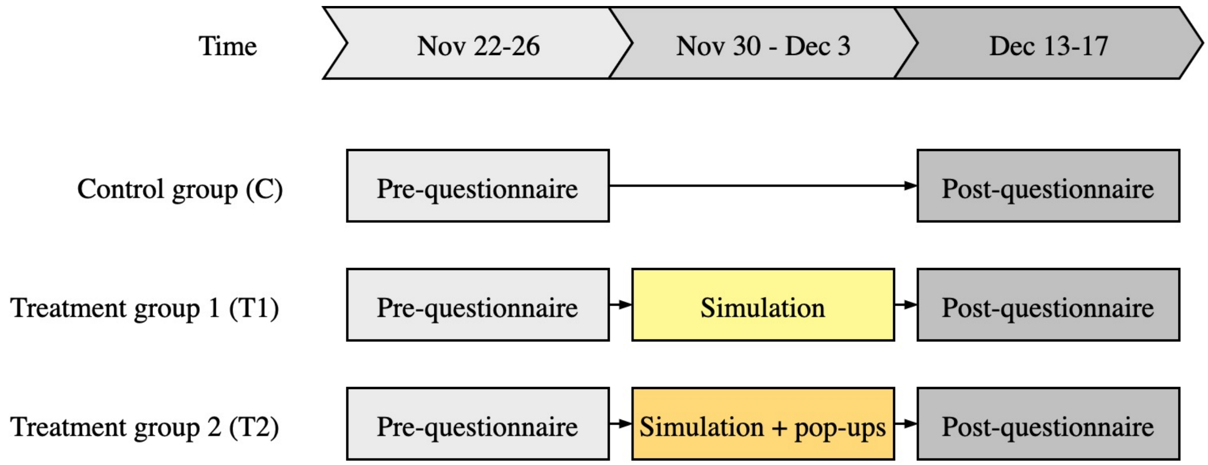

Figure 1. The experimental design and timeline

Of the 704 students who completed the pre-questionnaire, 234 were assigned to the control group (C), 235 to Treatment Group 1 (T1), and 235 to Treatment Group 2 (T2). Control participants received no additional activity between the two questionnaires. Students assigned to T1 and T2 were invited to complete an online experimental game between November 30 and December 3, 2023, designed to simulate a trading-app environment. In T1, the game was accompanied by basic instructions on its mechanics and by a set of short financial "pills"—concise summaries of fundamental investment concepts such as risk-return trade-offs and asset classes—presented before the start of the simulation. These pills constitute a baseline form of financial education, common to both treatment arms[2]. T2 followed the same game and schedule but additionally included in-game educational pop-ups delivered at pre-determined points during gameplay. These pop-ups drew attention to common behavioral pitfalls relevant to app-based trading—including diversification, overtrading, the disposition effect, availability bias, and herd behavior—without disclosing information about the game's payoff structure or providing personalized guidance (see Section 4 for details). This two-treatment design allows us to identify the effect of the simulation per se (T1 vs. C) and the incremental contribution of the behavioral prompts (T2 vs. T1). By design, both treatment arms rely on simulation-based learning rather than traditional classroom instruction. This choice reflects both the scalability objective of the study—the interventions are intended to be deployable asynchronously and at minimal cost—and sample size constraints that precluded the inclusion of an additional treatment arm with a conventional FE format.

[2]The same introductory instructions and financial pills were also provided to T2 participants; the two treatment arms differ exclusively in the presence of in-game pop-ups during the simulation (see Section 4 and Appendix C).

All participants completed a post-questionnaire between December 13 and 17, 2023. Each activity — whether a questionnaire or the simulation — had to be finished within a 12-hour window. The incentive scheme awarded a €5 reward for completing both questionnaires, supplemented by potential earnings from the simulation, and bank transfers were issued only upon successful completion of all required tasks. Overall, 488 students completed the full protocol (both questionnaires and, for treated students, the simulation) and constitute the completion sample used throughout the paper.[3] Because participation requirements differed across arms (treated students also had to complete the game), attrition is an important component of the study design. For transparency, we therefore document attrition patterns, assess whether dropout is systematically related to baseline measures, and derive worst-case bounds on treatment effects in Section 6. The structured timeline, randomized assignment, and the attrition analysis presented in Section 6 jointly support a causal interpretation of differences across arms for the completion sample.

Following a standard conceptualization in the FE literature, we operationalize financial decision-making competence along three complementary dimensions: financial knowledge, financial attitudes, and financial behaviors. This structure allows us to assess whether the intervention affects what participants know, what they report about risk, and how they allocate resources under uncertainty. This multidimensional perspective is supported by recent evidence that financial education can affect knowledge, attitudes, and psychological traits differentially, with substantial heterogeneity in individual responses (Gerrans et al., 2025). For financial attitudes and behaviors, we complement survey-based measures with incentivized choices observed in the trading simulation.

Financial knowledge refers to an understanding of fundamental financial concepts—such as interest rates, risk–return trade-offs, and the various types of financial instruments—as well as awareness of the challenges inherent in trading. We construct two knowledge indices. Our primary outcome is a core score based on nine standard items, following Ungeheuer and Weber (2021), which capture participants' grasp of key financial issues required for informed financial decisions. We additionally field five pop-up-specific questions and use them to compute an extended score ("score w/ biases") that combines the nine standard items with the five pop-up items. Unless explicitly stated otherwise, references to financial knowledge in the analyses below refer to the core nine-item score (the extended score is reported in the descriptive statistics).

[3]Two participants allocated their entire pre-test budget to the savings account, yielding an undefined baseline Sharpe ratio. The regression analyses that condition on this baseline measure therefore use an analytic sample of $N = 486$; details are provided in Section 5.1.

Financial attitudes are operationalized through stated risk attitudes, capturing how individuals report their willingness to bear risk in investment contexts. Stated risk attitudes are measured at both baseline and endline using the general willingness-to-take-risks item introduced by Dohmen et al. (2011), which Lönnqvist et al. (2015) identify as "the more adequate measure of individual risk attitudes for the analysis of behavior in economic (lab) experiments". Using the same item before and after the intervention allows us to assess whether FE affects reported risk tolerance in a standard and comparable format. To complement this survey-based measure, we also derive a behavioral measure of revealed risk-taking from participants' choices in the trading simulation. Specifically, we use the risky share, defined as the average fraction of available wealth allocated to risky assets during gameplay. The use of risky asset shares as a proxy for risk preferences is well established in both theoretical and empirical portfolio choice research: under a given investment opportunity set, the share invested in risky assets is monotonically related to risk aversion (Gomes and Michaelides, 2005), and has been widely used to infer individual risk attitudes from observed allocations (Calvet et al., 2009; Chiappori and Paiella, 2011; Guiso and Paiella, 2008). Because all participants face the same assets, payoffs, and information structure, variation in risky shares can be interpreted as reflecting differences in risk-taking within a common opportunity set, rather than differences in constraints or investment opportunities. This measurement strategy allows us to relate stated risk attitudes to revealed risk-taking in a controlled environment and to assess whether the intervention affects reported underlying risk preferences as well as actual allocation choices, following recent contributions that emphasize the value of integrating validated survey measures with incentivized behavioral data (Falk et al., 2023). By design, the simulation is more likely to affect how participants manage risk in a specific financial environment than to shift their deep underlying risk preferences.

Financial behaviors are assessed using both self-reported behaviors and actual behaviors observed during the experimental simulation. Self-reported financial behaviors are captured through a dedicated asset allocation question administered at both baseline and endline, following the approach of Cavezzali et al. (2015). In both waves, participants declare their intended allocation across the same set of assets featured in the simulation, based only on qualitative information about risk and return. Using these declared allocations, we compute comparable design-based Sharpe ratios before and after the intervention, providing an ex-ante measure of intended portfolio efficiency. During the simulation, actual trading behaviors—namely, portfolio allocations chosen across rounds—are recorded. Using the trading logs, we compute an in-game Sharpe ratio based on the participant's time-averaged portfolio weights across the rounds played. This yields a behavior-based measure of portfolio efficiency that is directly comparable to the questionnaire-based measure, as both rely on the same design-implied expected return vector μ and variance-covariance matrix Σ, derived from the game's known payoff structure rather than estimated from realized data (see Section4 for formal definitions), and differ only in the source of portfolio weights (declared versus observed during gameplay).

The research questions guiding our investigation are twofold. First, does participation in an experimental game that simulates a trading app environment improve participants' financial knowledge, attitudes, and behaviors compared to a control group receiving no intervention? Second, does the addition of in-game educational pop-up prompts—as provided exclusively to T2—lead to superior financial decision-making, particularly in terms of risk attitudes and portfolio efficiency, compared to providing only basic financial information (as in T1)?

4 The trading simulation

The trading simulation was designed to closely mimic the decision-making environment characteristic of modern trading apps, serving simultaneously as an educational intervention and a controlled laboratory for observing realistic investment behaviors. The game is fully implemented in oTree (Chen et al., 2016) and is structured into 20 rounds, with each round representing a separate allocation decision in a trading-app-like environment.

At the start of the game, each participant receives an initial endowment of 300 points (equivalent to €3). In each round, participants allocate their current point balance across the five investment options summarized in Table 1. The option set includes a zero-percent savings account (SA), which represents a risk-free alternative offering a guaranteed, neutral return, and four risky assets labeled A, B, C, and D. Each risky asset is characterized by a binary outcome mechanism: in each round it yields either a high or a low return, with outcomes generated by the program and communicated to the participant in real time. The underlying payoff probabilities (50-50 across the two scenarios) are fixed by design but not disclosed to participants, so that they are aware of uncertainty but cannot infer precise distributions or volatilities. This choice ensures that perceived risk relies on qualitative information rather than on calculable standard deviations, reflecting the incomplete information investors often face in app-based financial decisions. After returns are realized, the point balance updates accordingly. Point balances are bounded below at zero (limited liability): if returns would imply a negative balance, the participant's balance is set to zero and the game ends. Participants who exhaust their balance cannot proceed to subsequent rounds.

Table 1. Investment options

	Scenarios			
	S1	**S2**	**Risk**	**Return**
SA	0%	0%	Null	Null
Asset A	150%	-130%	Very high	Very high
Asset B	-40%	50%	Low	Low
Asset C	100%	-85%	High	High
Asset D	-20%	25%	Very low	Very low

To eliminate any potential framing effects or biases that might arise from associations with real-world securities, no actual asset names are used. Instead, assets are presented in abstract terms. It is worth noting that asset A includes a possible return of -130%, i.e., a loss that exceeds the initial amount invested. This feature is intentionally introduced to create an extreme downside scenario, ensuring that participants are confronted with the possibility of outcomes that would more than wipe out their investment in that asset. By doing so, the simulation provides a sharper contrast between safe and risky choices and allows us to examine decision-making under particularly adverse conditions.

Prior to the start of the game, participants are shown a graph that displays the historical returns of each asset, generated through a simulation of 15 hypothetical rounds for each asset. This historical performance chart is updated in real time as the game progresses. Notably, the presentation of historical returns is grounded in the practices of real trading apps, which routinely provide users with past performance data to inform their investment decisions. Together with the integration of quantitative payoff information and qualitative risk descriptions, this feature enhances the realism of the simulation by reproducing the informational environment typical of retail trading platforms, where investors observe partial performance histories but cannot infer precise return distributions. Participants can thus base their decisions on the same combination of partial, performance-based cues that characterize decision-making in real-world app environments.

At the conclusion of every round, after participants have made their allocation decisions, the game randomly selects one of two possible scenarios, labeled S1 and S2. In S1, assets A and C pay the high return while assets B and D pay the low return. Conversely, in S2, the outcomes are reversed so that assets A and C yield the low return and assets B and D yield the high return. This procedure establishes a precise correlation structure: assets A and C are perfectly positively correlated (they always pay the same scenario), as are assets B and D; by contrast, any cross-group pair (e.g., A and B, or C and D) is perfectly negatively correlated.

A naive $1/n$ allocation across all four risky assets therefore does not eliminate risk, because half the portfolio co-moves perfectly while the other half offsets it only partially.

This design serves two purposes. First, it mirrors the systematic co-movement patterns observed among real financial assets, where returns rarely behave independently. Second, it provides a sharp test of whether participants move beyond the naive-diversification heuristic documented by Huberman and Jiang (2006) and Benartzi and Thaler (2001), whereby investors distribute wealth evenly across available options—or the related "conditional $1/n$" rule of Shefrin and Statman (2000), which frames portfolios as collections of narrowly defined sub-portfolios. Because a $1/n$ split remains inefficient under this asset structure, genuine improvements in portfolio quality require participants to account for correlation, not merely to "hold many assets".

After each round, participants receive immediate graphical feedback that displays the evolution of their wealth over time. Figure 2 shows the screen seen by participants in T2. In this treatment arm, educational pop-ups were programmed to appear for the first time at fixed points during the simulation (rounds 1, 3, 5, 8, and 11). Each pop-up was displayed in a non-skippable window for 20 seconds, ensuring that participants who reached the relevant round were necessarily exposed to the content before continuing with the game. This feature limits differential take-up among exposed participants, while acknowledging that some participants may terminate early and therefore not reach later pop-ups. The diversification message, directly pertinent to the allocation task, was deliberately placed in round 1 to ensure that all participants encountered it. The four remaining topics were assigned to the subsequent pop-up slots through a single random draw conducted ex ante and then held constant across all participants. The rounds themselves were deliberately concentrated in the first half of the game to maximize the likelihood of exposure, given that some participants could exhaust their balance before completing all 20 rounds. Once displayed, each pop-up remained accessible for consultation in subsequent rounds, ensuring continuous availability throughout the game. This combined system of graphical feedback and educational prompts allowed participants to observe the impact of their allocation decisions and to adjust their strategies over time. Finally, to preserve incentives, final compensation is determined by a uniformly random draw of one round among those completed, with the selected round revealed only after the game, so that every allocation decision is payoff-relevant.

Figure 2. Graphic interface for T2

This comprehensive design, which integrates historical performance data and real-time feedback within a controlled simulation of trading app dynamics, provides a robust and realistic environment for examining the effects of our interventions on actual investment behaviors.

4.1 Measuring portfolio efficiency: A design-based Sharpe ratio

To quantify portfolio efficiency, we construct a Sharpe ratio that is fully determined by the experimental opportunity set. Expected returns and risk are not estimated from realized price paths but are derived from the game's payoff design. Let μ denote the vector of asset expected returns implied by the two equiprobable scenarios, and let Σ denote the corresponding variance-covariance matrix. The savings account is treated as a risk-free asset with zero return and zero variance by construction.

For any portfolio weight vector w, portfolio expected return and risk are defined as

$$E[R(w)] = w^\top \mu, \qquad \sigma(w) = \sqrt{w^\top \Sigma w},$$

and the design-based Sharpe ratio is given by

$$SR(w) = \frac{E[R(w)]}{\sigma(w)}.$$

For portfolios with $\sigma(w) = 0$, the Sharpe ratio is undefined. In our data this occurs for two participants who allocated their entire pre-test budget to the savings account.[4] In specifications that include baseline Sharpe ratio as a covariate, these observations are excluded. As a sensitivity check, we verified that replacing the baseline Sharpe ratio with baseline return and baseline volatility entered as separate controls retains the full completion sample of 488 students and yields qualitatively similar estimates across all specifications; these results are available upon request. Because all participants face the same μ and Σ, differences in $SR(w)$ reflect differences in allocation choices rather than differences in the return-generating process. The measure is therefore used as a common benchmark of risk-return efficiency implied by the experimental opportunity set, abstracting from path-dependent truncation due to limited liability.

We compute this measure in three settings. First, using the allocations declared in the pre-test questionnaire, we obtain a self-reported pre-test Sharpe ratio $SR(w_i^{pre})$. Second, using post-test declared allocations, we obtain $SR(w_i^{post})$. Third, using the trading logs, we compute an in-game Sharpe ratio based on the time-averaged portfolio weights across the rounds actually played. Specifically, if participant i completes T_i rounds (with $T_i \leq 20$ if the game ends early) and chooses weights w_{it} in round t, we define the average allocation as $\bar{w}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} w_{it}$ and set $SR_i^{game} = SR(\bar{w}_i)$.

This measure is intended to capture a participant's typical allocation profile under the common, design-implied risk-return structure, while abstracting from within-participant round-to-round variation.

[4] In principle, a zero-variance portfolio can also be constructed from perfectly negatively correlated risky assets (e.g., A and B in appropriate proportions), yielding a positive expected return. No participant in our sample chose such a combination.

5 Empirical strategy and results

Our empirical strategy employs an ordinary least squares (OLS) regression framework to identify the causal impact of the FE interventions delivered in T1 and T2 on a set of outcomes capturing financial knowledge, financial attitudes, and financial behaviors. Jointly estimating effects across all three dimensions—and complementing self-reported measures with incentivized choices from the trading simulation—allows for a more comprehensive assessment of the intervention than analyses that focus on a single outcome domain.

The completion sample consists of 488 students. For consistency across specifications, our main regression analyses use an analytic sample of N = 486, excluding two participants whose baseline design-based Sharpe ratio is undefined due to a zero-variance pre-test portfolio (100% savings account). For specifications focusing on in-game outcomes, the analytic sample correspondingly reduces to N = 306, as the same two observations—one from each treatment arm—are excluded by the same criterion. In our baseline specification, the regression model is given by

$$Y_i = \beta_0 + \beta_1 \cdot Treat_{1_i} + \beta_2 \cdot Treat_{2_i} + \lambda \cdot X_i + \epsilon_i,$$

where Y_i denotes the outcome of interest for student i, and $Treat_{1_i}$ and $Treat_{2_i}$ are indicators for assignment to treatment groups T1 and T2. The vector X_i includes a set of baseline characteristics measured prior to treatment assignment, such as pre-test financial knowledge, pre-test stated risk attitudes, pre-test portfolio performance indicators, and the time required to complete the pre-test questionnaire. These variables capture heterogeneity in initial conditions and improve the precision of the estimated treatment effects. In specifications focusing on in-game outcomes, we additionally incorporate information from the trading simulation logs, including measures of time spent reading instructions, average time per round, number of rounds played, and in-game earnings. These variables are post-treatment by construction and are included to characterize how participants interacted with the simulation environment, not as causal controls. Accordingly, specifications without process measures constitute the primary estimates throughout; augmented specifications that include them are reported for descriptive completeness and should not be given a causal interpretation.

5.1 Full-sample results

Descriptive statistics for the main variables—including financial knowledge scores, stated risk attitudes, and design-based portfolio measures—are reported in Appendix Tables13 and14. Pre-treatment comparability is high for financial knowledge and stated risk attitudes. We observe small baseline differences in implied pre-test returns (and, marginally, savings-account shares), which are controlled for in the regression specifications. Differences emerge primarily in post-test outcomes. Post-test financial knowledge increases markedly for treated students, with the largest gains in T2. While post-test stated risk attitudes remain stable across arms, portfolio-based outcomes reveal meaningful improvements in efficiency. In particular, volatility computed from post-test declared allocations declines for treated students, whereas differences in post-test returns are small and only marginally significant. As a result, post-test Sharpe ratios are higher in both treatment groups, with the largest gains again observed in T2.

5.1.1 Financial knowledge

Turning first to financial knowledge, Table 2 shows that both T1 and T2 significantly increase post-test knowledge scores. Relative to the control group mean of 0.57, the estimated coefficients imply gains of about 7-8% for students in T1 and about 15-16% for students in T2. Expressed in standard-deviation units of the pooled post-test score, these gains correspond to effect sizes of approximately 0.25-0.30 SD for T1 and about 0.5 SD for T2[5]. The estimates are precisely measured and remain stable when conditioning on pre-treatment covariates, including baseline knowledge, risk attitudes, and prior exposure to trading apps. A Wald test rejects the equality of the two treatment effects ($p = 0.008$), confirming that the pop-up-augmented simulation produces significantly larger knowledge gains than the baseline simulation.

[5]For context, a recent meta-meta-analysis covering 6,327 RCTs across health, finance, and sustainability domains reports an average bias-corrected treatment effect on behavior of 0.063 SD, with substantial heterogeneity (Kaiser et al., 2025). The T2 knowledge effect in our study substantially exceeds this benchmark, consistent with the meta-analytic finding that the most effective interventions combine active engagement with targeted educational content.

Table 2. Treatment effect on post-test knowledge

	Dependent variable:	
	Post-test score	
	(1)	(2)
T1	0.039**	0.048***
	(0.018)	(0.016)
T2	0.086***	0.089***
	(0.018)	(0.015)
Pre-test score		0.493***
		(0.045)
Pre-test time		0.005
		(0.010)
Pre-test trading app use_No, but		0.018
		(0.017)
Pre-test trading app use_Yes		0.030
		(0.020)
Pre-test stated risk attitude		−0.010**
		(0.004)
Pre-test SR		0.007
		(0.006)
Pre-test SA		−0.076**
		(0.031)
Observations	486	486
R^2	0.047	0.326
Adjusted R^2	0.043	0.313
Residual Std. Error	0.161 (df = 483)	0.137 (df = 476)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. Pre-test time and Pre-test SR enter in natural logarithms. The omitted category for trading app use is Pre-test trading app use_No (students who had never used a trading app and were unaware of what it was). T1: Treatment Group 1 (simulation with pills); T2: Treatment Group 2 (simulation with pills and pop-ups); SR: Sharpe ratio; SA: savings account. Wald test of H_0: $\beta_{T1} = \beta_{T2}$: $p = 0.009$ (column 1), $p = 0.008$ (column 2).

5.1.2 Financial attitudes

We next turn to financial attitudes, captured through a stated measure of risk attitudes and a behavioral measure of revealed risk-taking in the game. We begin with stated risk attitudes elicited via the Dohmen et al. (2011) item. Table 3 shows that neither T1 nor T2 has a statistically significant effect on post-test stated risk attitudes, and a Wald test confirms that the two treatment effects are themselves indistinguishable ($p = 0.328$).

We then examine revealed risk-taking observed during the trading simulation, as measured by the share of wealth allocated to risky assets. Because only treated students participated in the simulation, this analysis compares T2 to T1 rather than to the control group. Table 4 provides suggestive evidence that students in T2 allocate a higher fraction of their portfolio to risky assets than those in T1. In the specification with pre-treatment controls, the estimated coefficient is about +0.040, equivalent to roughly a 5% increase relative to the T1 mean of 0.75. This estimate is stable when in-game process measures—which capture how participants interacted with the simulation, such as time per round and number of rounds played—are added as descriptive controls, suggesting that the treatment difference is not driven by systematic differences in task engagement across arms. Overall, these results provide suggestive evidence that while stated risk attitudes remain unchanged, the augmented simulation affects students' willingness to take risk in payoff-relevant portfolio choices—a pattern that, when read together with the improvements in portfolio efficiency documented below, is consistent with improved understanding of portfolio risk rather than a shift in underlying risk tolerance.

Table 3. Treatment effect on post-test stated risk attitude

	Dependent variable:	
	Post-test stated risk attitude	
	(1)	(2)
T1	0.053	0.041
	(0.172)	(0.121)
T2	−0.130	−0.085
	(0.180)	(0.128)
Pre-test score		−0.792**
		(0.344)
Pre-test time		0.145*
		(0.080)
Pre-test trading app use_No, but		0.125
		(0.138)
Pre-test trading app use_Yes		0.445***
		(0.154)
Pre-test stated risk attitude		0.619***
		(0.035)
Pre-test SR		−0.132**
		(0.059)
Pre-test SA		−0.743***
		(0.271)
Observations	486	486
R^2	0.002	0.519
Adjusted R^2	−0.002	0.510
Residual Std. Error	1.599 (df = 483)	1.118 (df = 476)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. Pre-test time and Pre-test SR enter in natural logarithms. Omitted category and abbreviations as in Table 2. Wald test of $H_0\colon \beta_{T1} = \beta_{T2}$: $p = 0.310$ (column 1), $p = 0.328$ (column 2).

Table 4. Treatment effect on in-game revealed risk-taking

	Dependent variable:		
	In-game revealed risk-taking		
	(1)	(2)	(3)
T2	0.033	0.040*	0.042*
	(0.025)	(0.022)	(0.022)
Pre-test score		0.042	0.049
		(0.079)	(0.078)
Pre-test time		−0.022	−0.018
		(0.019)	(0.019)
Pre-test trading app use_No, but		0.080**	0.085***
		(0.030)	(0.030)
Pre-test trading app use_Yes		0.049	0.045
		(0.036)	(0.036)
Pre-test stated risk attitude		0.001	0.0001
		(0.007)	(0.007)
Pre-test SR		0.007	0.011
		(0.012)	(0.012)
Pre-test SA		−0.502***	−0.485***
		(0.061)	(0.061)
Time on instructions			0.001
			(0.008)
Time per round			−0.001
			(0.014)
Rounds played			−0.009***
			(0.002)
In-game earnings			0.002
			(0.003)
Observations	306	306	306
R^2	0.006	0.262	0.281
Adjusted R^2	0.003	0.243	0.252
Residual Std. Error	0.220 (df = 304)	0.192 (df = 297)	0.191 (df = 293)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. Pre-test time, Time on instructions, Time per round, and Pre-test SR enter in natural logarithms. Column (3) augments column (2) with in-game process measures; these are post-treatment descriptors of task engagement and should not be interpreted as causal mediators. Omitted category and abbreviations as in Table 2.

5.1.3 Financial behaviors

Finally, we turn to financial behaviors, captured through portfolio efficiency measures derived from both declared and realized allocations. Because the raw Sharpe ratio is right-skewed and bounded below, all regressions use its natural logarithm as the dependent variable. This transformation stabilizes the variance and yields residuals that more closely approximate normality; descriptive statistics for the untransformed Sharpe ratio are reported in Appendix B for completeness. We begin with self-reported behavior. Table 5 documents that both treatments significantly increase the Sharpe ratio computed from post-test declared allocations. Interpreting the coefficients from the specification with pre-treatment controls (column 2), the estimates imply an increase in the self-reported Sharpe ratio of about 27% for T1 and about 38% for T2 relative to the control group. While the T2 point estimate is larger, a Wald test does not reject equality of the two treatment effects ($p = 0.498$), indicating that the two interventions produce statistically comparable gains in declared portfolio efficiency. Baseline portfolio quality is a strong predictor of post-test efficiency, indicating persistence in allocation skills.

We then examine realized behavior during the trading simulation, focusing on differences between the two treatment groups. Table 6 shows that students in T2 achieve significantly higher in-game Sharpe ratios than those in T1. In the baseline specification without covariates, the estimated T2 coefficient is 0.411 log points, corresponding to an increase of about 51% relative to T1. Conditioning on pre-treatment characteristics yields a slightly smaller estimate of 0.367 log points (about 44%).

Augmenting the specification with in-game process measures characterizes how performance covaries with engagement during gameplay. The number of rounds played is strongly and positively associated with the in-game Sharpe ratio, and the treatment coefficient attenuates once these process measures are included. Consistent with our interpretation of these variables as descriptive (post-treatment) indicators of interaction with the task, this pattern indicates that performance differences co-vary with observed engagement during gameplay rather than providing a causal decomposition of treatment effects. Taken together, these results highlight that the augmented simulation not only improves intended portfolio choices, but also translates into more efficient behavior under repeated, payoff-relevant decision-making. Figure 3 provides a visual summary of the estimated treatment effects across all five outcome domains.

Table 5. Treatment effect on log post-test portfolio efficiency

	Dependent variable:	
	Log post-test SR	
	(1)	(2)
T1	0.218**	0.237**
	(0.109)	(0.101)
T2	0.359***	0.319***
	(0.118)	(0.108)
Pre-test score		0.609**
		(0.270)
Pre-test time		0.008
		(0.066)
Pre-test trading app use_No, but		0.256**
		(0.103)
Pre-test trading app use_Yes		0.102
		(0.111)
Pre-test stated risk attitude		−0.034
		(0.027)
Pre-test SR		0.452***
		(0.054)
Pre-test SA		−0.111
		(0.210)
Observations	486	486
R^2	0.020	0.214
Adjusted R^2	0.016	0.199
Residual Std. Error	1.062 (df = 483)	0.958 (df = 476)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. The dependent variable is the natural logarithm of the post-test design-based Sharpe ratio. Pre-test time, and Pre-test SR enter in natural logarithms. Omitted category and abbreviations as in Table 2. Wald test of $H_0\colon \beta_{T1} = \beta_{T2}$: $p = 0.291$ (column 1), $p = 0.498$ (column 2).

Table 6. Treatment effect on log in-game portfolio efficiency

	Dependent variable:		
	Log in-game SR		
	(1)	(2)	(3)
T2	0.411***	0.367***	0.312**
	(0.130)	(0.127)	(0.122)
Pre-test score		0.585	0.442
		(0.415)	(0.411)
Pre-test time		0.113	0.033
		(0.109)	(0.113)
Pre-test trading app use_No, but		0.224	0.180
		(0.180)	(0.177)
Pre-test trading app use_Yes		0.175	0.278
		(0.215)	(0.212)
Pre-test stated risk attitude		−0.015	−0.003
		(0.038)	(0.037)
Pre-test SR		0.248***	0.204***
		(0.060)	(0.058)
Pre-test SA		−0.118	−0.280
		(0.298)	(0.286)
Time on instructions			0.066
			(0.052)
Time per round			0.137
			(0.110)
Rounds played			0.091***
			(0.011)
In-game earnings			0.022
			(0.020)
Observations	306	306	306
R^2	0.032	0.098	0.189
Adjusted R^2	0.029	0.074	0.156
Residual Std. Error	1.138 (df = 304)	1.111 (df = 297)	1.061 (df = 293)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. The dependent variable is the natural logarithm of the in-game design-based Sharpe ratio. Pre-test time, Time on instructions, Time per round, and Pre-test SR enter in natural logarithms. Column (3) augments column (2) with in-game process measures; see note to Table 4. Omitted category and abbreviations as in Table 2.

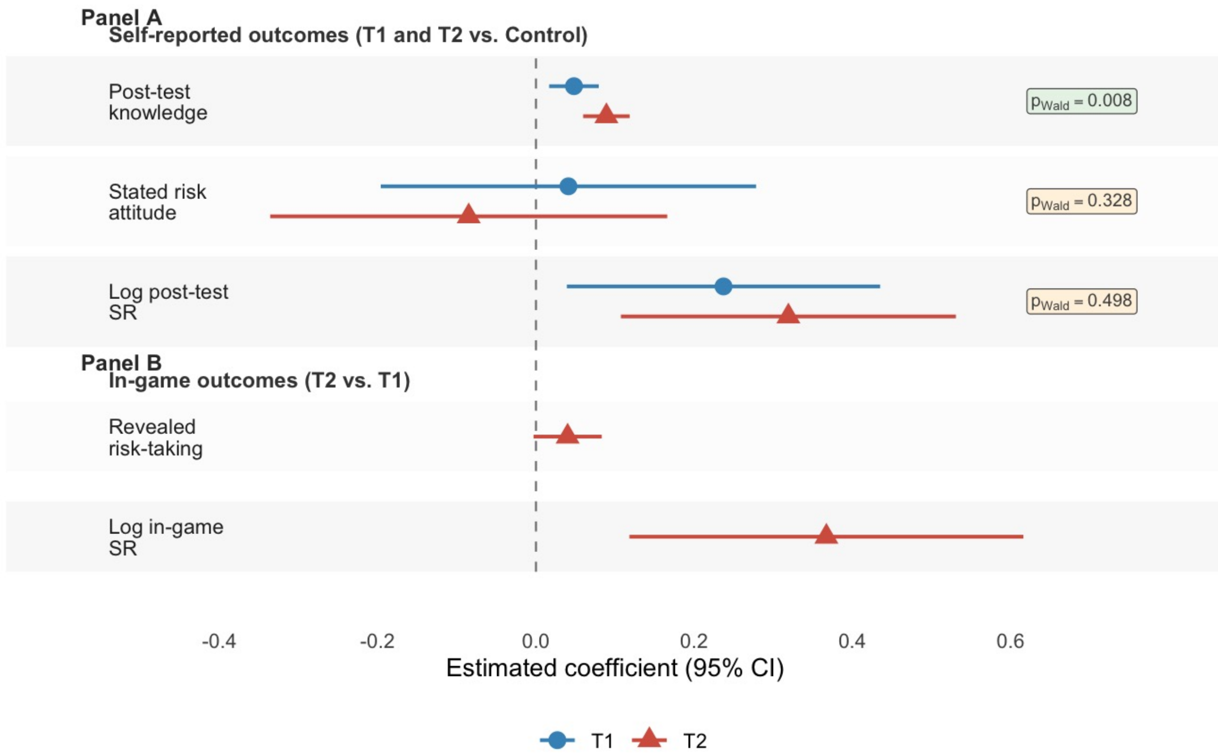

Figure 3. Treatment effects across outcomes

Note: Panel A reports estimated coefficients from specifications with the full set of pre-treatment controls for self-reported outcomes (column 2 of Tables 2, 3, and 5). Panel B reports the T2 coefficient from the corresponding in-game specifications (column 2 of Tables 4 and 6). Bars represent 95% confidence intervals based on HC1-robust standard errors. Wald test p-values for $H_0: \beta_{T1} = \beta_{T2}$ are reported in Panel A. Abbreviations as in Table 2.

5.2 Robustness with demographic controls and treatment effect heterogeneity

Analyses that leverage individual-level demographics and academic background are performed on the subset of participants who consented to data processing (319 students for post-test knowledge, stated risk attitude, and SR; 217 treated students for in-game outcomes). In addition to the baseline measures available for the full sample, the consented subsample includes information on gender, age, type of high school diploma, whether the student is enrolled in a major with above-median dropout rates, academic performance, and household socioeconomic status.

Academic performance is summarized by a weighted performance index (WPI), which combines achievement and study progress: $WPI_i = GPA_i \times \ln(1 + ECTS_i)$, where GPA_i

denotes the student's grade-point average on the Italian 18-30 scale, and $ECTS_i$ denotes the cumulative number of university credits earned by student i at the time of data collection[6]. The logarithmic transformation dampens the mechanical scale effect of credit accumulation, while preserving the interpretation of WPI as a composite proxy for performance and progress. Household socioeconomic status is proxied by the *Indicatore della Situazione Economica Equivalente* (ISEE), an Italian means-testing index that combines household income and wealth and is used to determine eligibility for university fee reductions and scholarships[7]. Variable definitions and descriptive statistics are reported in Appendix Tables 15 and 16.

Randomization remains broadly credible within the consented subsamples: age, gender composition, and WPI are comparable across arms. At the same time, we observe differences in the ISEE distribution and in some baseline portfolio measures (notably pre-test return and savings-account shares). These variables are therefore controlled for in the corresponding regression specifications.

When augmenting the post-test knowledge specification with student-level demographics (Table 7), the main pattern is unchanged. T1 yields a positive but modest gain in this subsample (about +0.16 SD), while T2 delivers a larger and precisely estimated improvement (about +0.45 SD). In the fully controlled specification, the estimated coefficients correspond to +0.038 for T1 and +0.074 for T2. Relative to the control-group mean in the consented sample (0.59), this implies increases of about 6-7% and about 12-13%, respectively. Demographic and background covariates do not display statistically significant associations with post-test knowledge in this specification, with the exception of WPI.

---

[6]We refer to the European Credit Transfer and Accumulation System (ECTS), in which credits quantify the workload required to achieve the learning outcomes of a course or programme. As a benchmark, 60 ECTS credits correspond to the workload of a full-time student during one academic year. Workload norms are often described as roughly 1,500-1,800 hours per year (about 25-30 hours per ECTS credit), although implementation can vary across systems. See European Commission (ECTS overview) and the ECTS Users' Guide.

[7]ISEE brackets are grouped into three categories: scholarship recipients, mid-to-low brackets (1-8), and the highest bracket (9). The omitted category in all regressions is ISEE_Max.

Table 7. Treatment effect on post-test knowledge (consented subsample)

	Dependent variable:	
	Post-test score	
	(1)	(2)
T1	0.024	0.038*
	(0.022)	(0.020)
T2	0.069***	0.074***
	(0.021)	(0.019)
Female		0.009
		(0.022)
Age		−0.00003
		(0.004)
WPI		0.001**
		(0.0003)
ISEE_Scholarship		−0.016
		(0.021)
ISEE_Mid-low		−0.002
		(0.017)
Non-scientific HS		−0.012
		(0.019)
High-dropout major		0.013
		(0.016)
Pre-test score		0.448***
		(0.052)
Pre-test time		0.013
		(0.013)
Pre-test trading app use_No, but		0.021
		(0.023)
Pre-test trading app use_Yes		0.016
		(0.029)
Pre-test stated risk attitude		−0.005
		(0.005)
Pre-test SR		0.008
		(0.008)
Pre-test SA		−0.100***
		(0.038)
Observations	319	319
R^2	0.033	0.321
Adjusted R^2	0.026	0.285
Residual Std. Error	0.158 (df = 316)	0.136 (df = 302)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. Pre-test time and Pre-test SR enter in natural logarithms. The omitted ISEE category is ISEE_Max (highest income bracket). Omitted category for trading app use and remaining abbreviations as in Table 2. Wald test of $H_0\colon \beta_{T1} = \beta_{T2}$: $p = 0.043$ (column 1), $p = 0.064$ (column 2).

Results for stated risk attitudes are consistent with the full-sample analysis. Table 8 shows no statistically significant effect of either treatment on post-test stated risk attitudes once demographics are included. By contrast, post-test stated risk attitudes display substantial persistence, with pre-test stated risk attitudes strongly predicting post-test values. Prior exposure to trading apps is positively associated with stated risk tolerance, while other demographic and socioeconomic controls do not meaningfully shift the pattern.

Turning to revealed risk-taking during the simulation, Table 9 indicates that students in T2 allocate a significantly larger share of wealth to risky assets than students in T1. The estimated effect is stable across specifications and amounts to about +0.05 in the model with demographic and baseline controls. Academic performance is positively related to risk-taking in the game, while students in high-dropout majors allocate a lower risky share. Baseline savings account allocations are strongly negatively associated with in-game revealed risk-taking, consistent with persistence in safe-versus-risky investment orientation. When adding in-game process measures, the treatment effect remains, and the number of rounds played is negatively related to the risky share, suggesting that participants who stay longer in the task may not necessarily do so by systematically increasing risk exposure.

Table 8. Treatment effect on post-test stated risk attitude (consented subsample)

	Dependent variable:	
	Post-test stated risk attitude	
	(1)	(2)
T1	0.010	0.013
	(0.217)	(0.176)
T2	−0.169	−0.093
	(0.215)	(0.170)
Female		−0.142
		(0.183)
Age		0.015
		(0.044)
WPI		0.002
		(0.002)
ISEE_Scholarship		0.176
		(0.190)
ISEE_Mid-low		−0.116
		(0.155)
Non-scientific HS		−0.103
		(0.159)
High-dropout major		−0.059
		(0.139)
Pre-test score		−0.662
		(0.434)
Pre-test time		0.253**
		(0.100)
Pre-test trading app use_No, but		0.254
		(0.182)
Pre-test trading app use_Yes		0.474**
		(0.210)
Pre-test stated risk attitude		0.575***
		(0.047)
Pre-test SR		−0.071
		(0.072)
Pre-test SA		−0.545
		(0.351)
Observations	319	319
R^2	0.003	0.474
Adjusted R^2	−0.003	0.446
Residual Std. Error	1.566 (df = 316)	1.164 (df = 302)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. Pre-test time and Pre-test SR enter in natural logarithms. Omitted ISEE category and abbreviations as in Table 7. Wald test of $H_0\colon \beta_{T1} = \beta_{T2}$: $p = 0.400$ (column 1), $p = 0.508$ (column 2).

Table 9. Treatment effect on in-game revealed risk-taking (consented subsample)

	Dependent variable:		
	In-game revealed risk-taking		
	(1)	(2)	(3)
T2	0.038	0.049^{*}	0.052^{**}
	(0.030)	(0.025)	(0.025)
Female		−0.023	−0.020
		(0.037)	(0.038)
Age		−0.006	−0.004
		(0.007)	(0.007)
WPI		0.001^{*}	0.001^{**}
		(0.001)	(0.001)
ISEE_Scholarship		0.007	0.007
		(0.038)	(0.037)
ISEE_Mid-low		0.017	0.019
		(0.029)	(0.029)
Non-scientific HS		0.019	0.016
		(0.035)	(0.034)
High-dropout major		-0.074^{***}	-0.071^{**}
		(0.028)	(0.029)
Pre-test score		0.003	0.010
		(0.089)	(0.087)
Pre-test time		−0.004	−0.002
		(0.020)	(0.020)
Pre-test trading app use_No, but		0.077^{**}	0.083^{**}
		(0.037)	(0.037)
Pre-test trading app use_Yes		0.061	0.058
		(0.045)	(0.045)
Pre-test stated risk attitude		−0.001	−0.001
		(0.008)	(0.008)
Pre-test SR		0.008	0.011
		(0.013)	(0.014)
Pre-test SA		-0.557^{***}	-0.546^{***}
		(0.069)	(0.071)
Time on instructions			−0.0001
			(0.009)
Time per round			−0.001
			(0.017)
Rounds played			-0.008^{***}
			(0.003)
In-game earnings			0.002
			(0.005)
Observations	217	217	217
R^2	0.007	0.349	0.363
Adjusted R^2	0.003	0.300	0.302
Residual Std. Error	0.221 (df = 215)	0.185 (df = 201)	0.185 (df = 197)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. Pre-test time, Time on instructions, Time per round, and Pre-test SR enter in natural logarithms. Column (3) augments column (2) with in-game process measures; these are post-treatment descriptors of task engagement and should not be interpreted as causal mediators. Omitted ISEE category and abbreviations as in Table 7.

For self-reported portfolio efficiency (Table 10), treatment effects remain positive and economically meaningful after controlling for student demographics. In the fully specified model, the estimated coefficients are +0.275 for T1 and +0.366 for T2, indicating that the augmented intervention continues to outperform the baseline simulation also within the consented subsample. Baseline portfolio efficiency remains a strong predictor of post-test SR, reinforcing the pattern of persistence observed in the full sample. By contrast, demographic covariates do not display systematic or robust associations with self-reported portfolio efficiency.

Table 10. Treatment effect on log post-test portfolio efficiency (consented subsample)

	Dependent variable:	
	Log post-test SR	
	(1)	(2)
T1	0.271*	0.275**
	(0.142)	(0.136)
T2	0.399***	0.366**
	(0.146)	(0.141)
Female		−0.198
		(0.138)
Age		−0.011
		(0.032)
WPI		−0.001
		(0.002)
ISEE_Scholarship		−0.005
		(0.151)
ISEE_Mid-low		0.016
		(0.134)
Non-scientific HS		0.174
		(0.131)
High-dropout major		−0.165
		(0.120)
Pre-test score		0.408
		(0.353)
Pre-test time		0.033
		(0.093)
Pre-test trading app use_No, but		0.276**
		(0.135)
Pre-test trading app use_Yes		0.124
		(0.162)
Pre-test stated risk attitude		−0.062*
		(0.036)
Pre-test SR		0.468***
		(0.074)
Pre-test SA		−0.305
		(0.271)
Observations	319	319
R^2	0.022	0.227
Adjusted R^2	0.016	0.186
Residual Std. Error	1.109 (df = 316)	1.009 (df = 302)

Note: *p<0.1; **p<0.05; ***p<0.01. The dependent variable is the natural logarithm of the post-test design-based Sharpe ratio. HC1-robust standard errors in parentheses. Pre-test time, and Pre-test SR enter in natural logarithms. Omitted ISEE category and abbreviations as in Table 7. Wald test of H_0: $\beta_{T1} = \beta_{T2}$: $p = 0.433$ (column 1), $p = 0.542$ (column 2).

For realized portfolio efficiency during the trading simulation (Table 11), the T2 premium persists and remains statistically significant across specifications. In the model with baseline and demographic controls, the estimated T2 coefficient is about +0.473, and it attenuates to roughly +0.429 when in-game process measures are added. Engagement remains strongly related to performance: the number of rounds completed is again a robust positive correlate of in-game SR, while most demographic controls are small and imprecisely estimated. As in the full sample, baseline portfolio efficiency predicts realized performance, suggesting that the intervention builds on pre-existing heterogeneity in allocation ability but does not depend on a specific demographic subgroup.

Table 11. Treatment effect on log in-game portfolio efficiency (consented subsample)

	Dependent variable:		
	Log in-game SR		
	(1)	(2)	(3)
T2	0.500***	0.473***	0.429***
	(0.159)	(0.159)	(0.152)
Female		−0.178	−0.189
		(0.194)	(0.189)
Age		0.005	−0.006
		(0.030)	(0.029)
WPI		−0.002	−0.002
		(0.003)	(0.003)
ISEE_Scholarship		0.059	0.071
		(0.228)	(0.218)
ISEE_Mid-low		−0.016	−0.070
		(0.183)	(0.175)
Non-scientific HS		−0.237	−0.188
		(0.179)	(0.169)
High-dropout major		−0.286*	−0.262*
		(0.161)	(0.154)
Pre-test score		0.484	0.299
		(0.504)	(0.493)
Pre-test time		0.068	0.007
		(0.127)	(0.136)
Pre-test trading app use_No, but		0.278	0.196
		(0.223)	(0.223)
Pre-test trading app use_Yes		0.013	0.110
		(0.283)	(0.285)
Pre-test stated risk attitude		−0.0001	0.005
		(0.050)	(0.048)
Pre-test SR		0.282***	0.234***
		(0.073)	(0.071)
Pre-test SA		−0.381	−0.550
		(0.366)	(0.350)
Time on instructions			0.059
			(0.060)
Time per round			0.211
			(0.130)
Rounds played			0.097***
			(0.015)
In-game earnings			0.026
			(0.027)
Observations	217	217	217
R^2	0.044	0.156	0.249
Adjusted R^2	0.039	0.093	0.177
Residual Std. Error	1.171 (df = 215)	1.138 (df = 201)	1.084 (df = 197)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. The dependent variable is the natural logarithm of the in-game design-based Sharpe ratio. Pre-test time, Time on instructions, Time per round, and Pre-test SR enter in natural logarithms. Column (3) augments column (2) with in-game process measures; see note to Table 9. Omitted ISEE category and abbreviations as in Table 7.

To further assess whether the intervention's effects vary across individuals, we explored treatment heterogeneity along three dimensions that have been consistently linked to financial learning and behavioral outcomes: gender, academic performance, and socioeconomic status (SES). These moderators are among the most frequently examined in empirical FE studies, reflecting persistent disparities in baseline financial literacy and potential differences in responsiveness to instruction.

Existing evidence suggests that gender gaps in financial literacy emerge early and can persist despite instruction, with females often showing lower baseline knowledge and weaker behavioral improvements (Kalwij et al., 2019; Sconti, 2022; Torma et al., 2023). Academic performance is also commonly associated with learning and retention, although moderation by achievement is not consistently observed across settings (Becchetti et al., 2013; Kalmi and Rahko, 2022; Torma et al., 2023). Finally, SES—proxied by parental resources such as education or income—may shape both exposure to financial concepts and opportunities to apply them; accordingly, prior evaluations report mixed evidence, with some finding larger gains among low-SES participants and others finding muted or null heterogeneity (Amagir et al., 2022; Bover et al., 2024; Iterbeke et al., 2022; Sconti, 2022). Building on this literature, we focus on a parsimonious set of interactions between treatment assignment and three pre-treatment characteristics—female, WPI, and ISEE—to test whether T1 and T2 yield systematically different impacts across groups.

Across outcomes capturing financial knowledge, financial attitudes, and financial behaviors, we do not find systematic or robust heterogeneity in treatment effects. Interaction estimates are generally small, do not follow a consistent pattern across measures, and are in some cases sensitive to specification—particularly for in-game outcomes, where the inclusion of interaction terms introduces notable instability in the main treatment coefficients, reflecting collinearity in a relatively small sample ($N = 217$). While isolated interaction terms reach marginal significance for in-game portfolio efficiency—specifically, WPI$\times$T2 and ISEE mid-low$\times$T2 are each significant at the 10% level—neither result is part of a consistent pattern across outcomes or moderators. Given the sample sizes involved, the study is underpowered to detect all but large moderating effects. The absence of significant heterogeneity should therefore be interpreted as consistent with, but not conclusive evidence for, homogeneous treatment effects across the student population. In particular, we do not find evidence that individual demographic or academic characteristics systematically moderate the estimated average treatment effects. The full set of heterogeneity specifications, including detailed tables and discussion, is reported in AppendixD.

6 Differential attrition and internal validity

As noted in Section 3, treated students faced higher participation requirements than control students, making differential attrition a structurally relevant concern. This section documents dropout patterns, assesses whether attrition is systematically related to pre-treatment characteristics, and derives worst-case bounds on treatment effects under arbitrary selection on unobservables.

6.1 Dropout patterns and baseline balance among completers

Table 12 summarizes attrition across the three experimental arms. Panel A shows that dropout rates amount to 23.1% in the control group, 34.5% in T1, and 34.5% in T2. A Pearson chi-squared test rejects equality of dropout rates across groups ($p = 0.009$), confirming that the additional participation requirement imposed on treated students translated into systematically higher attrition relative to the control group. Crucially, however, dropout rates are identical across the two treatment arms, which implies that differential attrition cannot confound comparisons between T1 and T2.

To assess whether attrition is systematically related to pre-treatment financial competence, Panel B of Table 12 compares baseline financial knowledge between completers and dropouts. The two groups are statistically indistinguishable (Welch t-test, $p = 0.48$), indicating that students who abandoned the study were not systematically less knowledgeable than those who completed it. This result is corroborated by the multivariate logit model in Panel C, where dropout is regressed on treatment assignment and pre-test financial knowledge: while assignment to T1 and T2 significantly raises the probability of dropout, baseline knowledge is not a significant predictor of attrition. Together, these results suggest that the excess dropout in the treatment arms reflects the additional effort required by the simulation task rather than systematic self-selection based on financial ability.

Table 12. Attrition analysis

	C	T1	T2
Panel A: Dropout rates			
Number of students (pre-test)	234	235	235
Number of dropouts	54	81	81
Dropout rate (%)	23.1	34.5	34.5
Pearson's χ^2 test (p-value)		0.009	
Panel B: Baseline financial knowledge			
Pre-test score (non-dropouts)		0.577 (0.159)	
Pre-test score (dropouts)		0.585 (0.130)	
Welch two-sample t-test (p-value)		0.475	
Panel C: Logit model for dropout			
T1		0.563^{***}	
		(0.208)	
T2		0.560^{***}	
		(0.208)	
Pre-test score		0.360	
		(0.510)	
Observations		704	

Note: ***p<0.01. Panel A: Pearson chi-squared test of equal dropout rates across arms. Panel B: means; standard deviations in parentheses. Panel C: logit regression for dropout (= 1); HC3-robust standard errors in parentheses. Abbreviations as in Table 2.

6.2 Bounding treatment effects under worst-case selection on unobservables

The absence of selection on observed baseline competence mitigates, but does not eliminate, concerns about differential attrition. Even if completers and dropouts are comparable in terms of measured pre-treatment knowledge, selection on unobserved dimensions—such as motivation, engagement, or interest in financial topics—cannot be ruled out. To formally assess the sensitivity of the main estimates to this possibility, we compute Lee (2009) bounds on the treatment effects for the two primary outcomes available across all three experimental arms: post-test financial knowledge and post-test SR.

The Lee (2009) bounding approach constructs a worst-case interval for the average treatment effect by trimming, from the group with the higher completion rate, the fraction of its observations that accounts for the excess completion relative to the other arm. Because Lee bounds are nonparametric and do not condition on baseline covariates, completion rates are computed on the full completion sample of 488 students rather than on the analytic sample of 486 used in the regression analyses, which excludes two participants whose baseline Sharpe ratio is undefined. The control group exhibits a completion rate of 76.9%, compared with 65.5% in either treatment group; the resulting trimming fraction is 14.8%.[8] The lower bound is obtained by trimming the most favorable control-group observations (those with the highest outcome values), thereby maximizing the counterfactual performance of the control group; the upper bound is obtained by trimming the least favorable ones. Under the monotonicity assumption—that treatment assignment can only increase or only decrease the probability of completing the study—this procedure recovers a sharp identified set for the average treatment effect on compliers. Monotonicity is credible in this context: the simulation task constitutes an additional participation requirement that can only raise the cost of completion, and therefore can only increase, never decrease, the probability of dropout. Confidence intervals for each bound are obtained via 2,000 bootstrap replications and are reported at the 90% level, consistent with standard practice in applications of this method.

The bounds are reported in Figure 4. For each treatment-outcome combination, the thick segment represents the identified interval from the lower to the upper bound; the thin segments extending beyond it represent the 90% bootstrap confidence intervals on each bound; and the diamond with its associated confidence interval represents the OLS point estimate from the specification without controls estimated on the same completion sample of 488 students. The dashed horizontal line marks zero.

[8]The trimming fraction equals the difference in completion rates divided by the completion rate of the group being trimmed: $(0.769 - 0.655)/0.769 \approx 0.148$.

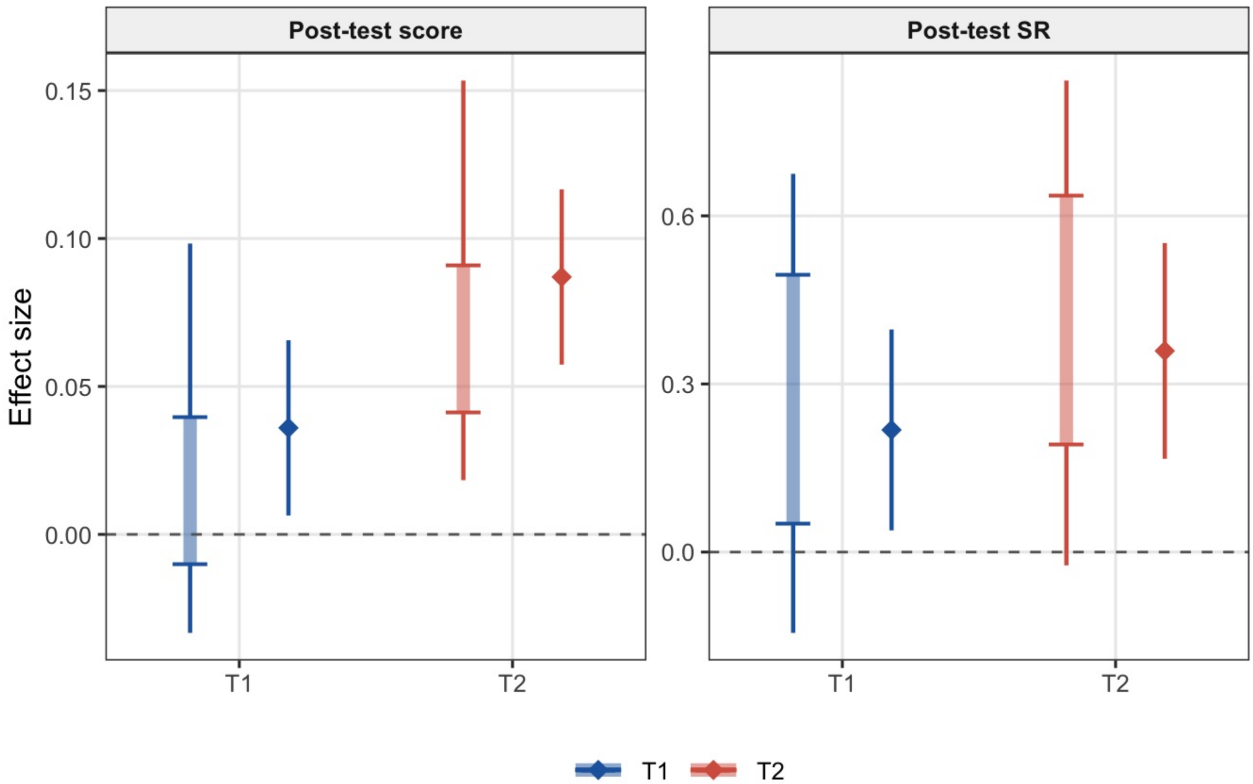

Figure 4. Lee (2009) bounds on treatment effects

Note: Thick segments: identified interval [lower bound, upper bound]. Thin segments: 90% bootstrap confidence intervals on each bound (2,000 replications). Diamond with thin segment: OLS point estimate with 90% confidence interval from the specification without pre-treatment controls estimated on the full completion sample ($N = 488$), ensuring that the OLS reference and the Lee bounds are computed on the same population and target the same unconditional estimand. The main regression estimates with the full set of pre-treatment controls on the analytic sample ($N = 486$) are reported in column 2 of Tables 2 and 5. Dashed line: zero. Trimming fraction: 14.8% in all four cases. Post-test SR refers to the self-reported Sharpe ratio computed from declared post-test allocations and enters in natural logarithms, as in the main specifications. Lee bounds for the in-game Sharpe ratio are not reported: the control group did not participate in the simulation by design, so the outcome is structurally missing rather than subject to attrition; for the T2 vs. T1 comparison, dropout rates are identical across arms, so the bounds collapse to the point estimate.

The results are informative and consistent with the broader pattern of evidence in the paper. For T2, the identified interval is strictly positive for both outcomes. For post-test knowledge, the lower bound is +0.041 (90% CI: $[+0.018, +0.088]$) and the upper bound is +0.091 (90% CI: $[+0.061, +0.153]$), with the confidence interval on the lower bound lying entirely above zero. This indicates that the effect of the augmented simulation on financial knowledge is robust to arbitrary selection on unobservables under monotonicity. For log post-test SR, both bounds are positive as well: the lower bound is +0.192 (90% CI: $[-0.024, +0.416]$) and the upper bound is +0.637 (90% CI: $[+0.395, +0.842]$). While the point estimates of the identified interval are strictly above zero, the confidence interval on the lower bound marginally includes zero, indicating that the effect on declared portfolio efficiency, though consistently positive across the identified set, cannot be statistically distinguished from zero at the 90% level under the most adverse selection scenario.

For T1, the picture is more nuanced. The bounds for log post-test SR are both positive (+0.051 and +0.496), confirming that the effect of the baseline simulation on declared portfolio efficiency is positive across the identified set, although the confidence interval on the lower bound extends below zero (90% CI: $[-0.144, +0.267]$). By contrast, the bounds for post-test knowledge straddle zero: the lower bound is -0.010 (90% CI: $[-0.033, +0.038]$), indicating that the smallest estimated effect in the paper—the gain in financial knowledge attributable to the simulation alone—cannot be signed with certainty under the most adverse selection scenario. This result is consistent with the relatively modest magnitude of the T1 knowledge effect reported in Table 2 and does not affect comparisons between T1 and T2, for which attrition is balanced by construction.

An important scope condition applies to this bounding exercise. Because the control group did not participate in the simulation by design, the in-game Sharpe ratio—which captures realized portfolio efficiency from actual, incentivized allocation decisions—is structurally unavailable for the control arm and therefore cannot be subjected to Lee bounds. For the T2 vs. T1 comparison, where the in-game SR is observed in both arms, dropout rates are identical (34.5%), so the bounds collapse to the OLS point estimate and no trimming is required. The robustness evidence presented here thus pertains to self-reported outcomes (financial knowledge and declared portfolio efficiency); the in-game results reported in Section 5.1 are instead protected from differential attrition by the balanced dropout rates across treatment arms.

Taken together, the evidence in this section supports the causal interpretation of the main findings. The effect of the augmented simulation on financial knowledge is the most robust result: both bounds and the associated confidence intervals are strictly positive. For self-reported portfolio efficiency, the identified intervals are positive for both treatments, although inference on the lower bounds is less precise. The pattern reinforces a central finding of the paper: the pop-up-augmented intervention not only produces larger average effects, but generates improvements that remain credible under conservative assumptions about selective study completion. The baseline simulation yields more modest and less precisely bounded effects, consistent with its smaller estimated treatment effects in the main analysis.

7 Conclusion

This paper studies the effects of two short, simulation-based FE interventions that differ in the degree of behavioral scaffolding and user guidance embedded in the learning environment. Using a randomized design with 704 university students, we show that both interventions improve students' financial knowledge and declared portfolio choices relative to a no-intervention control group, while the augmented simulation generates significantly larger gains in financial knowledge and consistently larger point estimates across all other outcomes. A key appeal of these interventions lies in their simplicity: they are short, asynchronous, non-personalized, and therefore easily scalable. Yet, they provide meaningful effects weeks after the intervention—and even in general domains of financial knowledge that were not fully or even explicitly covered in the simulation game.

In terms of financial knowledge, both treatment arms produce statistically significant gains relative to the control group, with effect sizes of approximately 0.25-0.30 SD for the baseline simulation and about 0.5 SD for the augmented version. These results confirm that even brief, app-like interventions can deliver meaningful learning gains in financial knowledge. Stated risk attitudes, by contrast, are unaffected by either intervention: neither treatment shifts stated risk tolerance as measured by the standard survey item, and the two treatment effects are themselves indistinguishable.

The strongest and most robust effects emerge for portfolio efficiency. Both treatments substantially increase the Sharpe ratio computed from declared post-test allocations—a self-reported measure—relative to the control group. When focusing on realized behavior during the incentivized trading game, students exposed to the augmented simulation achieve markedly higher in-game Sharpe ratios than those in the baseline treatment. In parallel, revealed risk-taking—measured by the share of wealth allocated to risky assets during the simulation—is higher in the augmented group, suggesting that T2 participants not only diversify more efficiently but also allocate a larger fraction of their budget to risky assets. Descriptive specifications that condition on post-treatment process measures show that these treatment differences attenuate once task engagement is accounted for; because these measures are themselves affected by treatment, this pattern describes co-variation between engagement and performance rather than identifying a causal mediation channel.

The joint increase in risky-asset shares and portfolio efficiency, combined with the absence of any change in stated risk tolerance, points to a specific mechanism. Rather than shifting participants' underlying appetite for risk, the pop-up prompts appear to enhance their understanding of the risk-return trade-off—particularly the role of asset correlation in determining portfolio risk. Participants who receive the prompts take on more risk and manage it more efficiently, consistent with improved comprehension of diversification rather than a simple increase in risk-seeking. This interpretation is supported by evidence that behavior in incentivized and context-rich environments is more responsive to changes in decision-making frameworks and informational cues than are self-reported attitudes (Chapman et al., 2025; Falk et al., 2023). This mechanism is directly relevant to the policy concerns motivating the study. European households continue to allocate a disproportionate share of their savings to bank deposits, while the rapid growth of trading apps exposes new investors to complex allocation decisions they are often ill-prepared to handle. The finding that brief, embedded prompts can improve how users manage the risk-return trade-off—without requiring changes in underlying risk preferences—suggests a practical pathway for scalable interventions within these platforms.

Overall, the findings contribute to the FE literature in three ways. First, they provide causal evidence that a fast, light-touch trading simulation can generate sizable gains in financial knowledge and portfolio efficiency. Second, they show that short, behaviorally informed prompts embedded in a realistic decision environment can further improve realized investment performance, even in the absence of changes in stated risk preferences. Third, by combining self-reported intentions with realized, incentivized behavior under a common return-generating structure, the study offers a tighter link between financial competence and observed portfolio decision-making than is typical in FE evaluations.

Several limitations point to directions for future research. External validity is bounded by a single institutional setting and a relatively homogeneous sample of students from a single STEM university. Within this sample, treatment effect heterogeneity analyses by gender, socioeconomic status, and academic performance do not reveal systematic moderation of the main effects, although the study is underpowered to detect anything but large differential impacts. Outcomes are measured over a short horizon, preventing conclusions about persistence over time. Because treated students complete the simulation in addition to both questionnaires, the estimated effects relative to the control group capture the combined impact of the intervention package—including any additional engagement with financial content; the T2 versus T1 comparison, which holds total task structure approximately constant, isolates the incremental contribution of the pop-ups more cleanly. Attrition differs across experimental arms due to differential participation requirements. However, baseline competence does not predict dropout and worst-case bounding confirms that the T2 effect on financial knowledge—the most precisely bounded self-reported outcome—is robust to arbitrary selection on unobservables. Future work should examine longer-run effects, replicate the intervention in more diverse populations, and investigate mechanisms more directly, including how attention and engagement interact with behavioral prompts over time. Enriching the simulation with additional realistic features—such as aggregate shocks, time-varying correlations, or transaction costs—would further test whether the competence gains documented here extend to more complex decision environments.

Our findings are consistent with the perspective that well-designed, behaviorally informed simulations can do more than transmit financial concepts: they can meaningfully improve how novice investors allocate risk in environments that increasingly shape everyday financial decision-making. Fast-paced, easily accessible trading apps have become an established part of the financial landscape; our results suggest that financial education can adapt to this environment without sacrificing scalability.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT (OpenAI) and Claude (Anthropic) to support language editing, improve clarity and coherence, assist with manuscript formatting, and refine selected passages of the text. The tools were not used to generate data, conduct statistical analyses, or produce empirical results. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Abreu, M. and Mendes, V. (2010). Financial literacy and portfolio diversification. *Quantitative Finance*, 10(5):515–528.

Amagir, A., Groot, W., Maassen van den Brink, H., and Wilschut, A. (2018). A review of financial-literacy education programs for children and adolescents. *Citizenship, Social and Economics Education*, 17(1):56–80.

Amagir, A., van den Brink, H. M., Groot, W., and Wilschut, A. (2022). Savewise: The impact of a real-life financial education program for ninth grade students in the netherlands. *Journal of Behavioral and Experimental Finance*, 33:100605.

Amari, M., Salhi, B., and Jarboui, A. (2020). Evaluating the effects of sociodemographic characteristics and financial education on saving behavior. *International Journal of Sociology and Social Policy*, 40(11/12):1423–1438.

Ambuehl, S., Bernheim, B. D., and Lusardi, A. (2014). The effect of financial education on the quality of decision making. NBER Working Paper 20618, National Bureau of Economic Research.

Aren, S. and Zengin, A. N. (2016). Influence of financial literacy and risk perception on choice of investment. *Procedia - Social and Behavioral Sciences*, 235:656–663.

Aucejo, E. M. and Wong, K. (2025). The effect of feedback on student performance. *Journal of Public Economics*, 241:105274.

Bajo, E., Barbi, M., and Sandri, S. (2015). Financial literacy, households' investment behavior, and risk propensity. *Journal of Financial Management, Markets and Institutions*, 3(1):157–174.

Barber, B. M., Huang, X., Odean, T., and Schwarz, C. (2022). Attention-induced trading and returns: Evidence from robinhood users. *The Journal of Finance*, 77(6):3141–3190.

Bayar, Y., Sezgin, H. F., Öztürk, Ö. F., and Şaşmaz, M. Ü. (2020). Financial literacy and financial risk tolerance of individual investors: Multinomial logistic regression approach. *SAGE Open*, 10(3):2158244020945717.

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

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5, Part 2):9–49.

Benartzi, S. and Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, 91(1):79–98.

Blake, M., Shah, A., and Astorino, L. (2022). The future of capital markets: Democratization of retail investing. Insight report, World Economic Forum.

Bover, O., Hospido, L., and Villanueva, E. (2024). The impact of high school financial education on financial knowledge and saving choices: Evidence from a randomized trial in spain. *Journal of Human Resources*.

Brade, R., Himmler, O., and Jäckle, R. (2022). Relative performance feedback and the effects of being above average—field experiment and replication. *Economics of Education Review*, 89:102268.

Brescia Morra, C., Colonnello, D., Gargantini, M., Sandrelli, G., and Trovatore, G. (2025). La gamification degli investimenti finanziari. Technical Report 32, CONSOB, Roma.

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

Carpena, F., Cole, S., Shapiro, J., and Zia, B. (2019). The abcs of financial education: Experimental evidence on attitudes, behavior, and cognitive biases. *Management Science*, 65(1):346–369.

Carpena, F. and Zia, B. (2020). The causal mechanism of financial education: Evidence from mediation analysis. *Journal of Economic Behavior & Organization*, 177:143–184.

Cavezzali, E., Gardenal, G., and Rigoni, U. (2015). Risk taking behaviour and diversification strategies: Do financial literacy and financial education play a role? *Journal of Financial Management, Markets and Institutions*, 3(1):121–156.

Chapkovski, P., Khapko, M., and Zoican, M. (2025). Gamified risk-taking. *Journal of Behavioral and Experimental Finance*, 46:101049.

Chapkovski, P., Khapko, M., and Zoican, M. (2026). Trading gamification and investor behavior. *Management Science*, 72(1):32–56.

Chapman, J., Ortoleva, P., Snowberg, E., Yariv, L., and Camerer, C. (2025). Reassessing qualitative self-assessments and experimental validation. Technical report, National Bureau of Economic Research.

Chaudhry, S. and Kulkarni, C. (2021). Design patterns of investing apps and their effects on investing behaviors. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, DIS '21, pages 777–788, New York, NY, USA. Association for Computing Machinery.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Chiappori, P.-A. and Paiella, M. (2011). Relative risk aversion is constant: Evidence from panel data. *Journal of the European Economic Association*, 9(6):1021–1052.

Dinç Aydemir, S. and Aren, S. (2017). Do the effects of individual factors on financial risk-taking behavior diversify with financial literacy? *Kybernetes*, 46(10):1706–1734.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Duraj, K., Grunow, D., Haliassos, M., Laudenbach, C., and Siegel, S. (2025). Rethinking the stock market participation puzzle: A qualitative approach. Technical report, CESifo Working Paper.

European Fund and Asset Management Association (EFAMA) (2025). Asset management in europe: An overview of the asset management industry. Technical report, EFAMA. 17th edition.

Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*, 69(4):1935–1950.

Gerrans, P., Hoffmann, A. O. I., McNair, S. J., and Pallant, J. I. (2025). More than objective knowledge: Exploring heterogeneity in individuals' response to a financial education initiative across multiple financial literacy domains. *Pacific-Basin Finance Journal*, 90:102669.

Gomes, F. and Michaelides, A. (2005). Optimal life-cycle asset allocation: Understanding the empirical evidence. *The Journal of Finance*, 60(2):869–904.

Guiso, L. and Jappelli, T. (2009). Financial literacy and portfolio diversification. CSEF Working Papers 212, Centre for Studies in Economics and Finance (CSEF), University of Naples.

Guiso, L. and Paiella, M. (2008). Risk aversion, wealth, and background risk. *Journal of the European Economic Association*, 6(6):1109–1150.

Hermansson, C. and Jonsson, S. (2021). The impact of financial literacy and financial interest on risk tolerance. *Journal of Behavioral and Experimental Finance*, 29:100450.

Horn, S., Jamison, J. C., Karlan, D., and Zinman, J. (2023). Five-year impacts of group-based financial education and savings promotion for ugandan youth. *Review of Economics and Statistics*, pages 1–53.

Huberman, G. and Jiang, W. (2006). Offering versus choice in 401(k) plans: Equity exposure and number of funds. *The Journal of Finance*, 61(2):763–801.

Iterbeke, K., Schelfhout, W., and De Witte, K. (2022). The role of students' interests during computer-assisted learning: A field experiment. *Computers in Human Behavior*, 130:107168.

Kaiser, T., Kloidt, J., Mata, J., and Hertwig, R. (2025). A meta-meta-analysis of behavior change interventions: Two tales of behavior change. Technical report, CESifo Working Paper.

Kaiser, T. and Menkhoff, L. (2020). Financial education in schools: A meta-analysis of experimental studies. *Economics of Education Review*, 78:101930.

Kalmi, P. and Rahko, J. (2022). The effects of game-based financial education: New survey evidence from lower-secondary school students in finland. *The Journal of Economic Education*, 53(2):109–125.

Kalwij, A., Alessie, R., Dinkova, M., Schonewille, G., Van der Schors, A., and Van der Werf, M. (2019). The effects of financial education on financial literacy and savings behavior: Evidence from a controlled field experiment in dutch primary schools. *Journal of Consumer Affairs*, 53(3):699–730.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Linciano, N., Costa, D., Gentile, M., and Soccorso, P. (2019). Report on financial investments of italian households: Behavioural attitudes and approaches. 2019 survey. Technical report, CONSOB.

Litterscheidt, R. and Streich, D. J. (2020). Financial education and digital asset management: What's in the black box? *Journal of Behavioral and Experimental Economics*, 87:101573.

Lobão, J. (2020). What to do in financial markets? preferences and incoherences of future investors. *Corvinus Journal of Sociology and Social Policy*, 11(2):3–22.

Lönnqvist, J.-E., Verkasalo, M., Walkowitz, G., and Wichardt, P. C. (2015). Measuring individual risk attitudes in the lab: Task or ask? an empirical comparison. *Journal of Economic Behavior & Organization*, 119:254–266.

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

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

Mouna, A. and Jarboui, A. (2015). Financial literacy and portfolio diversification: An observation from the tunisian stock market. *International Journal of Bank Marketing*, 33(6):808–822.

Muñoz-Murillo, M., Álvarez-Franco, P. B., and Restrepo-Tobón, D. A. (2020). The role of cognitive abilities on financial literacy: New experimental evidence. *Journal of Behavioral and Experimental Economics*, 84:101482.

Muralidharan, K., Singh, A., and Ganimian, A. J. (2019). Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review*, 109(4):1426–1460.

Oberrauch, L. and Kaiser, T. (2024). Financial education or incentivizing learning-by-doing? evidence from an rct with undergraduate students. *Journal of Behavioral and Experi-*

mental Finance, 43:100954.

Pitthan, F. and De Witte, K. (2025). How learning about behavioural biases can improve financial literacy? *International Review of Economics & Finance*, 99:103989.

Sconti, A. (2022). Digital vs. in-person financial education: What works best for generation z? *Journal of Economic Behavior & Organization*, 194:300–318.

Shefrin, H. and Statman, M. (2000). Behavioral portfolio theory. *Journal of Financial and Quantitative Analysis*, 35(2):127–151.

Torma, J., Barbić, D., and Ivanov, M. (2023). Analyzing the effects of financial education on financial literacy and financial behaviour: A randomized field experiment in croatia. *The South East European Journal of Economics and Business*, 18(2):63–86.

Ungeheuer, M. and Weber, M. (2021). The perception of dependence, investment decisions, and stock prices. *The Journal of Finance*, 76(2):797–844.

Van Klaveren, C., Vonk, S., and Cornelisz, I. (2017). The effect of adaptive versus static practicing on student learning: Evidence from a randomized field experiment. *Economics of Education Review*, 58:175–187.

West, T. and Cull, M. (2020). Future expectations and financial satisfaction. *Economic Papers: A Journal of Applied Economics and Policy*, 39(4):318–335.

Yoong, J. (2011). Financial illiteracy and stock market participation: Evidence from the rand american life panel. In Mitchell, O. S. and Lusardi, A., editors, *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*, pages 76–97. Oxford University Press.

A Transparency and design choices

To facilitate transparency and replicability in lieu of a pre-registration, this appendix documents and discusses details regarding the choices made by the researchers in the experimental design and outcome measures. Attrition patterns and their implications for the causal interpretation of the main estimates are discussed in Section 6.

A.1 General set-up of the experiment

The study evaluates two FE interventions against a no-intervention control group: a baseline simulation game and an augmented version that integrates educational pop-ups into the same simulation. As is standard when comparing a basic intervention to a more comprehensive one, the prior expectation is that both interventions produce positive effects, with the augmented version yielding larger gains. This expectation follows from the design rationale: since the two treatments are identical in terms of scalability and cost, and nearly identical in duration, the pop-ups were introduced precisely because they were expected to strengthen the educational channel of the simulation.

A.2 Outcomes and specifications

In terms of outcomes, we distinguish three outcome domains: financial knowledge, financial attitudes, and financial behaviors. Financial knowledge is captured through test scores, while financial behaviors are captured through both self-reported and incentivized portfolio-based measures. Intended behavior is measured via a portfolio allocation task that mirrors the structure of the experimental game. Based on declared allocations, we compute a design-based Sharpe ratio that reflects ex-ante portfolio efficiency, with expected returns and risk imputed from the game's known payoff structure rather than estimated from data. Behavior-based outcomes are derived from actual allocation choices made during the simulation. Using these observed in-game allocations, we compute the same design-based Sharpe ratio, relying on the identical return vector and variance–covariance matrix. As described in Section 3, self-reported and in-game Sharpe ratios are therefore directly comparable and differ only in the source of portfolio weights. All measures are well established and follow the standard of the FE literature (Cavezzali et al. 2015; Dohmen et al. 2011; Ungeheuer and Weber 2021; see Section 3 for a detailed discussion). The contribution of this paper lies within the design and evaluation of the intervention, not the choice of outcome measure. As self-reported and incentivized outcomes are available for comparisons between distinct groups, we do not specify one main outcome category: for comparisons with the control group, we have to rely on self-reported outcomes, as the incentivized outcomes are not available for the control group by definition. However, whenever we compare the two different treatment groups, we can use the incentivized measures in addition, since both groups played the payoff-relevant simulation game. The choice of outcome variables follows directly from the experimental design. As an exploratory extension, we also examine in-game revealed risk-taking, measured by the share of funds allocated to risky assets during the simulation.

In terms of covariates, we include pre-treatment characteristics that are standard in the literature and improve precision (e.g., baseline outcomes and prior exposure to trading apps). For transparency, each table reports specifications both without controls and with the full set of pre-treatment controls.

Variable definitions and descriptive statistics for the full and consented samples are reported in Appendix B. Section 5.2 discusses treatment effect heterogeneity along selected individual characteristics. These analyses are explicitly exploratory since they go beyond the simple comparison of treatment group averages, and for some of these heterogeneities there is no strong ex-ante hypothesis for why the treatment(s) should have stronger effects for certain subgroups of students. The full set of heterogeneity specifications is reported in Appendix D.

B Variable definitions and descriptive statistics

This appendix reports variable definitions and descriptive statistics for the full sample and for the consented subsample.

Table 13. Variable definitions (full sample)

Variable	Description	Value/range
Individual characteristics		
Pre-test (Post-test) trading app use_No	Whether the student had never used a trading app and did not know what it was before (after) the intervention	1; otherwise
Pre-test (Post-test) trading app use_No, but	Whether the student had never used a trading app but knew what it was before (after) the intervention	1; otherwise
Pre-test (Post-test) trading app use_Yes	Whether the student had used a trading app before (after) the intervention	1; otherwise
Financial knowledge		
Pre-test score	Core financial knowledge score (9 standard items) measured before the intervention	from 0.11 to 0.88
Post-test score	Core financial knowledge score (9 standard items) measured after the intervention	from 0.11 to 1.00
Pre-test score (w/ biases)	Extended knowledge score (14 items): core score plus 5 pop-up questions	from 0.14 to 1
Post-test score (w/ biases)	Extended knowledge score (14 items): core score plus 5 pop-up questions	from 0.14 to 0.93
Financial attitudes		
Pre-test (Post-test) stated risk attitude	The student's self-reported risk attitude before (after) the intervention	from 2 to 10
In-game revealed risk-taking (risky share)	Average share of the portfolio allocated to risky assets during the simulation	from 0.1425 to 1
Financial behaviors		
Pre-test (Post-test) return	Implied expected portfolio return from the pre-test (post-test) allocation	from 0.0025 to 0.10
Pre-test (Post-test) volatility	Implied portfolio standard deviation from the pre-test (post-test) allocation	from 0.0025 to 1.40
Pre-test SR	Design-based Sharpe ratio implied by the pre-test declared asset allocation	from 0.0714 to 7
Post-test SR	Design-based Sharpe ratio implied by the post-test declared asset allocation	from 0.0714 to 21
In-game SR	Design-based Sharpe ratio computed from time-averaged in-game portfolio weights	from 0.0768 to 9.91
Pre-test (Post-test) SA	Percentage of the portfolio allocated to the savings account in the pre-test (post-test)	from 0 to 1.00
Experimental game data		
Pre-test time	Time (in seconds) the student took to complete the pre-test questionnaire	from 180 to 19,920
Time on instructions	Time (in seconds) spent reading the game instructions	from 5 to 3,227
Time per round	Average time (in seconds) spent per round during the simulation	from 8.90 to 775.15
Rounds played	Total number of rounds completed during the simulation	from 1 to 20
In-game earnings	Total earnings obtained during the simulation (in euros)	from 0 to 23.02

Table 14. Descriptive statistics (full sample)

Variable	**C**	**T1**	**T2**	**p-value**
Pre-test score (w/ biases)	0.61 (0.15)	0.61 (0.16)	0.61 (0.14)	>0.9
Pre-test score	0.58 (0.16)	0.57 (0.17)	0.58 (0.15)	0.9
Post-test score (w/ biases)	0.60 (0.15)	0.61 (0.14)	0.70 (0.14)	<0.001
Post-test score	0.57 (0.17)	0.61 (0.17)	0.66 (0.15)	<0.001
Pre-test stated risk attitude	5.89 (1.66)	5.92 (1.77)	5.88 (1.66)	>0.9
Post-test stated risk attitude	6.04 (1.64)	6.08 (1.50)	5.88 (1.69)	0.5
Pre-test return	0.048 (0.014)	0.047 (0.016)	0.045 (0.015)	0.037
Pre-test volatility	0.23 (0.19)	0.25 (0.22)	0.21 (0.19)	0.3
Pre-test SR	0.56 (1.10)	0.55 (1.05)	0.57 (1.03)	0.5
NA	0	1	1	
Post-test return	0.049 (0.013)	0.046 (0.015)	0.046 (0.015)	0.093
Post-test volatility	0.26 (0.20)	0.23 (0.20)	0.22 (0.23)	0.006
Post-test SR	0.43 (0.86)	0.80 (2.04)	1.14 (2.89)	0.084
Pre-test SA	0.20 (0.20)	0.23 (0.20)	0.25 (0.21)	0.051
Post-test SA	0.20 (0.19)	0.24 (0.22)	0.22 (0.21)	0.3
Pre-test time (s)	1,916 (2,794)	1,412 (926)	1,606 (1,580)	0.5
Pre-test trading app use				0.4
No	40 (22%)	30 (19%)	24 (16%)	
No, but	91 (51%)	83 (54%)	93 (60%)	
Yes	49 (27%)	41 (27%)	37 (24%)	
Post-test trading app use				<0.001
No	34 (19%)	5 (3.2%)	7 (4.5%)	
No, but	100 (56%)	96 (62%)	105 (68%)	
Yes	46 (26%)	53 (34%)	42 (27%)	
In-game revealed risk-taking (risky share)	–	0.75 (0.23)	0.79 (0.21)	0.2
In-game SR	–	0.71 (1.17)	1.29 (2.29)	0.002
Time on instructions (s)	–	425 (463)	503 (643)	0.8
Time per round (s)	–	83 (90)	91 (76)	0.028
Rounds played	–	18.7 (3.7)	19.2 (3.3)	0.13
In-game earnings (€)	–	4.3 (3.2)	4.7 (4.0)	0.7
N	180	154	154	

Note: Continuous variables: means with standard deviations in parentheses. Categorical variables: frequencies and percentages. Post-test variables highlighted in light gray. *p*-values: Kruskal-Wallis tests for continuous variables (Wilcoxon for variables below the dashed line); Pearson chi-squared for categorical variables. Variables below the dashed line are computed only for treated students. SR: Sharpe ratio; SA: savings account. "No, but" corresponds to "no, but I know what it is"; "No" to "no, and I have no idea what it is".

Table 15. Variable definitions (consented sample)

Variable	Description	Value/range
Individual characteristics		
Female	Whether the student is female	1; otherwise
Age	The student's age in years	from 18 to 37
ISEE_Scholarship	Whether the student received a scholarship in the past academic year	1; otherwise
ISEE_Mid-low	Whether the student's ISEE falls within brackets 1–8	1; otherwise
ISEE_Max	Whether the student's ISEE is in the highest (ninth) bracket	1; otherwise
Non-scientific high school	Whether the student obtained a diploma from a non-scientific high school	1; otherwise
High-dropout major	Whether the student is enrolled in a major whose dropout rate exceeds the median	1; otherwise
Pre-test (Post-test) trading app use_No	Whether the student had never used a trading app and did not know what it was before (after) the intervention	1; otherwise
Pre-test (Post-test) trading app use_No, but	Whether the student had never used a trading app but knew what it was before (after) the intervention	1; otherwise
Pre-test (Post-test) trading app use_Yes	Whether the student had used a trading app before (after) the intervention	1; otherwise
Academic performance		
WPI	An index combining GPA and university credits: $WPI = GPA \times \ln(1 + ECTS)$	from 26.34 to 150.19
Financial knowledge		
Pre-test score	The student's financial knowledge score before the intervention	from 0.11 to 0.88
Post-test score	The student's financial knowledge score after the intervention	from 0.11 to 1.00
Financial attitudes		
Pre-test (Post-test) stated risk attitude	The student's self-reported risk attitude before (after) the intervention	from 2 to 10
In-game revealed risk-taking (risky share)	Average share of the portfolio allocated to risky assets during the simulation	from 0.15 to 1
Financial behaviors		
Pre-test (Post-test) return	Implied expected portfolio return from the pre-test (post-test) allocation	from 0.00625 to 0.10
Pre-test (Post-test) volatility	Implied portfolio standard deviation from the pre-test (post-test) allocation	from 0.0025 to 1.40
Pre-test SR	Design-based Sharpe ratio implied by the pre-test declared asset allocation	from 0.0714 to 6.20
Post-test SR	Design-based Sharpe ratio implied by the post-test declared asset allocation	from 0.0714 to 21.00
In-game SR	Design-based Sharpe ratio computed from time-averaged in-game portfolio weights	from 0.0807 to 13.26
Pre-test (Post-test) SA	Percentage of the portfolio allocated to the savings account in the pre-test (post-test)	from 0 to 0.9
Experimental game data		
Pre-test time	Time (in seconds) the student took to complete the pre-test questionnaire	from 180 to 19,920
Time on instructions	Time (in seconds) spent reading the game instructions	from 5 to 3,227
Time per round	Average time (in seconds) spent per round during the simulation	from 12.65 to 506.80
Rounds played	Total number of rounds completed during the simulation	from 2 to 20
In-game earnings	Total earnings obtained during the simulation (in euros)	from 0 to 21.43

Table 16. Descriptive statistics (consented subsample)

Variable	**C**	**T1**	**T2**	**p-value**
Female	0.25 (0.43)	0.27 (0.44)	0.28 (0.45)	0.9
Age	20.42 (1.48)	20.79 (2.29)	20.58 (1.57)	0.7
WPI	103 (26)	106 (25)	105 (25)	0.7
ISEE				0.023
Scholarship	36 (35%)	22 (21%)	23 (21%)	
Mid-low	33 (32%)	37 (35%)	52 (46%)	
Max	33 (32%)	46 (44%)	37 (33%)	
Non-scientific high school	0.31 (0.47)	0.21 (0.41)	0.24 (0.43)	0.2
High-dropout major	0.44 (0.50)	0.39 (0.49)	0.36 (0.48)	0.5
Pre-test score	0.59 (0.15)	0.56 (0.18)	0.58 (0.16)	0.5
Post-test score	0.59 (0.15)	0.62 (0.16)	0.66 (0.16)	0.007
Pre-test stated risk attitude	5.76 (1.64)	5.81 (1.76)	5.76 (1.64)	>0.9
Post-test stated risk attitude	5.99 (1.58)	6.00 (1.55)	5.82 (1.57)	0.5
Pre-test return	0.049 (0.014)	0.046 (0.014)	0.044 (0.015)	0.010
Pre-test volatility	0.24 (0.17)	0.23 (0.18)	0.21 (0.20)	0.3
Pre-test SR	0.44 (0.76)	0.58 (1.12)	0.59 (1.11)	0.9
Post-test return	0.051 (0.014)	0.046 (0.014)	0.045 (0.014)	0.014
Post-test volatility	0.28 (0.21)	0.22 (0.19)	0.20 (0.19)	0.005
Post-test SR	0.45 (0.87)	0.90 (2.35)	1.23 (3.19)	0.10
Pre-test SA	0.19 (0.20)	0.24 (0.20)	0.27 (0.21)	0.006
Post-test SA	0.17 (0.17)	0.24 (0.21)	0.23 (0.22)	0.046
Pre-test time (s)	1,864 (2,990)	1,401 (976)	1,676 (1,812)	0.5
Pre-test trading app use				0.7
No	23 (23%)	21 (20%)	21 (19%)	
No, but	52 (51%)	54 (51%)	66 (59%)	
Yes	27 (26%)	30 (29%)	25 (22%)	
Post-test trading app use				0.002
No	16 (16%)	3 (2.9%)	5 (4.5%)	
No, but	60 (59%)	64 (61%)	78 (70%)	
Yes	26 (25%)	38 (36%)	29 (26%)	
In-game revealed risk-taking (risky share)	–	0.74 (0.23)	0.78 (0.22)	0.2
In-game SR	–	0.80 (1.30)	1.45 (2.48)	<0.001
Time on instructions (s)	–	456 (500)	499 (648)	0.4
Time per round (s)	–	82 (74)	92 (82)	0.13
Rounds played	–	18.79 (3.56)	19.27 (3.16)	0.15
In-game earnings (€)	–	4.3 (2.9)	4.5 (3.6)	>0.9
N	102	105	112	

Note: Continuous variables: means with standard deviations in parentheses. Categorical variables: frequencies and percentages. Post-test variables highlighted in light gray. *p*-values: Kruskal-Wallis tests for continuous variables (Wilcoxon for variables below the dashed line); Pearson chi-squared for categorical variables. Variables below the dashed line are computed only for treated students. ISEE proxies socioeconomic status via parental income brackets; the omitted category is ISEE_Max (highest bracket). SR: Sharpe ratio; SA: savings account. "No, but" corresponds to "no, but I know what it is"; "No" to "no, and I have no idea what it is".

C Pop-up messages

This appendix reports the exact wording of the five educational pop-up messages shown to participants in T2, in the order of their first appearance during the simulation. The diversification pop-up was fixed in round 1; the assignment of the four remaining topics to rounds 3, 5, 8, and 11 was determined by a single random draw and held constant across all participants. Each pop-up was displayed in a non-skippable window for 20 seconds and remained accessible in subsequent rounds once first shown. Since the intervention was administered in Italian, Figure 5 reproduces the original diversification pop-up, while Table 17 reports the full English wording of all messages for consistency with the language of the manuscript.

Figure 5. Original in-game diversification pop-up (Italian version)

Table 17. Educational pop-up messages shown in T2

Round	Topic	Message
1	Diversification	"Are you risk-averse? Risk-neutral? Or perhaps a risk-seeker? Regardless of your risk profile, portfolio diversification is a key strategy. Why? Because diversifying—i.e., investing in multiple assets—offers a major advantage: it reduces volatility and stabilizes your returns. Be careful though: one security is not the same as another. Trading platforms offer a wide range of options, often organized by popularity or performance, but 'different' in finance does not simply mean that asset A is not asset B. Rather, it means that assets are imperfectly correlated or move in opposite directions. A truly diversified portfolio is composed of assets that do not all follow the same path."
3	Overtrading	"Have you heard of 'overtrading'? It consists of investing excessively, compulsively, and irrationally—often making frequent and unnecessary adjustments to your portfolio in an attempt to earn gains as quickly as possible. Trading platforms, in particular, create a fertile environment for this risky behavior by allowing you to invest at any time."
5	Disposition effect	"Speaking of emotional and short-sighted decisions, meet the disposition effect. This is the tendency of investors—especially beginners—to sell assets that have gained value too quickly and hold on too long to those that have lost value, neglecting the importance of building a sustainable, long-term investment portfolio."
8	Availability bias	"Learning from the past is essential, but it is equally important to look to the future—especially when making investment decisions. The tendency to overestimate what has already happened, expecting for example that an asset's past return is a reliable indicator of future performance, is known as 'availability bias'. The use of animations and flashy graphics when profits are recorded makes trading platforms fertile ground for hasty decisions."
11	Herding effect	"Have you ever trusted someone and copied their behavior? When this tendency spreads on a large scale, we call it the 'herding effect'. In finance, this phenomenon can entail significant risks, leading to sudden crashes in the value of invested assets. Many trading platforms allow you to follow and freely imitate top investors, making a complete—and potentially dangerous—outsourcing of decisions seem appealing. Gaining knowledge and understanding your risk profile are essential; investing is not a game to be taken lightly."

D Treatment effect heterogeneity

This appendix reports the full set of treatment effect heterogeneity specifications discussed in Section 5.2. For each of the five primary outcomes—post-test financial knowledge, post-test stated risk attitude, in-game revealed risk-taking, post-test SR, and in-game SR—we estimate augmented models that include pairwise interactions between treatment indicators and three pre-specified moderators: gender (Female), academic performance (WPI), and socioeconomic status (ISEE). Each table presents three specifications, one per moderator, with the full set of baseline and demographic controls included throughout. All specifications are estimated on the consented subsample; for post-test outcomes (Tables 18, 19, 21), the sample consists of 319 students across all three arms, while for in-game outcomes (Tables 20 and 22), the sample consists of 217 treated students (T1 and T2). As noted in Section 5.2, these analyses are exploratory, and the sample sizes involved—particularly for in-game outcomes—afford limited power to detect interaction effects.

The dominant pattern across all five outcomes and three moderators is one of null or negligible heterogeneity. Across the full set of treatment-moderator interactions, only two reach marginal significance at the 10% level. No single moderator produces a consistent pattern of differential effects across outcomes, and no outcome displays significant heterogeneity along more than one dimension.

The two marginally significant interactions both emerge for the in-game Sharpe ratio (Table 22). WPI$\times$T2 is estimated at -0.012 ($p < 0.10$), suggesting that the pop-up treatment may yield smaller efficiency gains for students with higher academic performance. ISEE mid-low$\times$T2 is -0.646 ($p < 0.10$), pointing to a similar attenuation for students from mid-to-low socioeconomic backgrounds. Both estimates, however, should be interpreted with considerable caution. The in-game sample is small (N = 217), and the inclusion of interaction terms introduces substantial instability in the main treatment coefficients: the T2 estimate rises from 0.379 in the gender specification to 1.721 in the WPI specification, a pattern indicative of collinearity between the moderator and the interaction term rather than a genuine shift in the underlying treatment effect. This instability is visible, to a lesser degree, in the in-game revealed risk-taking specifications as well (Table 20), where the T2 coefficient moves from 0.031 to 0.178 when the WPI interaction is added.

Beyond the in-game outcomes, interaction estimates are uniformly small and far from significance. For post-test financial knowledge (Table 18), the gender interactions are negative but modest, while WPI and ISEE interactions are negligible. For post-test stated risk attitude (Table 19), no interaction term approaches significance, consistent with the null main effects on stated risk preferences documented in the main analysis and confirming that this null is not masking offsetting heterogeneity across subgroups. For post-test portfolio efficiency (Table

21), the gender interactions are again negative for both treatments and economically non-negligible—directionally consistent with the literature documenting gender gaps in financial literacy gains—but neither reaches statistical significance given the available sample size.

Taken together, the heterogeneity analysis does not reveal systematic or robust moderation of treatment effects by gender, academic performance, or socioeconomic status. The improvements documented in the main analysis appear to be broadly shared across the student population rather than concentrated in specific demographic or academic subgroups. At the same time, the study is underpowered to detect all but large moderating effects, particularly for in-game outcomes. The absence of significant heterogeneity should therefore be interpreted as consistent with, but not conclusive evidence for, homogeneous treatment effects.

Table 18. Treatment effect heterogeneity: post-test knowledge (consented subsample)

	Dependent variable:		
	Post-test score		
	(1)	(2)	(3)
	Gender	WPI	ISEE
T1	0.046**	−0.011	0.047
	(0.022)	(0.082)	(0.032)
T2	0.086***	0.101	0.077**
	(0.022)	(0.081)	(0.033)
Female	0.036	0.009	0.009
	(0.033)	(0.020)	(0.020)
Age	0.00003	0.00004	0.001
	(0.004)	(0.004)	(0.004)
WPI	0.001**	0.001	0.001**
	(0.0003)	(0.001)	(0.0003)
ISEE_Scholarship	−0.015	−0.014	−0.014
	(0.021)	(0.021)	(0.034)
ISEE_Mid-low	−0.003	−0.0002	0.009
	(0.018)	(0.018)	(0.034)
Non-scientific HS	−0.012	−0.014	−0.016
	(0.019)	(0.019)	(0.019)
High-dropout major	0.013	0.014	0.014
	(0.017)	(0.017)	(0.017)
Pre-test score	0.446***	0.445***	0.444***
	(0.051)	(0.051)	(0.051)
Pre-test time	0.012	0.014	0.014
	(0.013)	(0.013)	(0.013)
Pre-test trading app use_No, but	0.022	0.019	0.020
	(0.021)	(0.021)	(0.021)
Pre-test trading app use_Yes	0.017	0.015	0.017
	(0.026)	(0.026)	(0.026)
Pre-test SR	0.007	0.008	0.008
	(0.009)	(0.009)	(0.009)
Pre-test stated risk attitude	−0.005	−0.005	−0.005
	(0.005)	(0.005)	(0.005)
Pre-test SA	−0.099**	−0.096**	−0.102**
	(0.041)	(0.041)	(0.041)
Female × T1	−0.033		
	(0.044)		
Female × T2	−0.044		
	(0.043)		
WPI × T1		0.0005	
		(0.001)	
WPI × T2		−0.0003	
		(0.001)	
ISEE_Scholarship × T1			−0.026
			(0.050)
ISEE_Scholarship × T2			0.023
			(0.050)
ISEE_Mid-low × T1			−0.012
			(0.047)
ISEE_Mid-low × T2			−0.019
			(0.045)
Observations	319	319	319
R^2	0.323	0.323	0.324
Adjusted R^2	0.283	0.282	0.279
Residual Std. Error	0.136 (df = 300)	0.136 (df = 300)	0.136 (df = 298)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. Pre-test time and Pre-test SR enter in natural logarithms. Each column augments the fully controlled specification (Table 7, column 2) with pairwise interactions between treatment indicators and a single moderator. Column headers indicate the moderator. The omitted ISEE category is ISEE_Max (highest income bracket). Omitted category for trading app use and remaining abbreviations as in Table 2.

Table 19. Treatment effect heterogeneity: post-test stated risk attitude (consented subsample)

	Dependent variable:		
	Post-test stated risk attitude		
	(1)	(2)	(3)
	Gender	WPI	ISEE
T1	−0.067	0.164	0.032
	(0.192)	(0.703)	(0.272)
T2	−0.097	0.200	−0.282
	(0.192)	(0.698)	(0.285)
Female	−0.260	−0.138	−0.144
	(0.286)	(0.174)	(0.174)
Age	0.012	0.015	0.012
	(0.038)	(0.038)	(0.038)
WPI	0.002	0.004	0.002
	(0.003)	(0.005)	(0.003)
ISEE_Scholarship	0.180	0.177	0.088
	(0.180)	(0.180)	(0.290)
ISEE_Mid-low	−0.111	−0.114	−0.181
	(0.155)	(0.156)	(0.295)
Non-scientific HS	−0.105	−0.113	−0.102
	(0.159)	(0.161)	(0.163)
High-dropout major	−0.048	−0.050	−0.065
	(0.142)	(0.143)	(0.142)
Pre-test score	−0.631	−0.672	−0.654
	(0.434)	(0.434)	(0.435)
Pre-test time	0.260**	0.257**	0.252**
	(0.114)	(0.114)	(0.114)
Pre-test trading app use_No, but	0.249	0.242	0.250
	(0.179)	(0.182)	(0.180)
Pre-test trading app use_Yes	0.459**	0.468**	0.476**
	(0.220)	(0.220)	(0.221)
Pre-test SR	−0.069	−0.069	−0.064
	(0.076)	(0.076)	(0.077)
Pre-test stated risk attitude	0.578***	0.575***	0.579***
	(0.042)	(0.042)	(0.043)
Pre-test SA	−0.560	−0.541	−0.541
	(0.349)	(0.352)	(0.350)
Female × T1	0.324		
	(0.380)		
Female × T2	0.034		
	(0.369)		
WPI × T1		−0.001	
		(0.007)	
WPI × T2		−0.003	
		(0.007)	
ISEE_Scholarship × T1			0.077
			(0.425)
ISEE_Scholarship × T2			0.221
			(0.426)
ISEE_Mid-low × T1			−0.130
			(0.401)
ISEE_Mid-low × T2			0.300
			(0.390)
Observations	319	319	319
R^2	0.475	0.474	0.476
Adjusted R^2	0.444	0.442	0.441
Residual Std. Error	1.166 (df = 300)	1.168 (df = 300)	1.169 (df = 298)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. Pre-test time and Pre-test SR enter in natural logarithms. Each column augments the fully controlled specification (Table 8, column 2) with pairwise interactions between treatment indicators and a single moderator. Column headers indicate the moderator. Omitted ISEE category and abbreviations as in Table 18.

Table 20. Treatment effect heterogeneity: in-game revealed risk-taking (consented subsample)

	Dependent variable:		
	In-game revealed risk-taking		
	(1)	(2)	(3)
	Gender	WPI	ISEE
T2	0.031	0.178	0.066
	(0.030)	(0.118)	(0.042)
Female	−0.059	−0.024	−0.023
	(0.046)	(0.034)	(0.034)
Age	−0.005	−0.006	−0.005
	(0.007)	(0.007)	(0.007)
WPI	0.001*	0.002**	0.001**
	(0.001)	(0.001)	(0.001)
ISEE_Scholarship	0.004	0.010	−0.008
	(0.037)	(0.037)	(0.051)
ISEE_Mid-low	0.016	0.021	0.046
	(0.029)	(0.030)	(0.042)
Non-scientific HS	0.020	0.014	0.017
	(0.033)	(0.033)	(0.033)
High-dropout major	−0.078***	−0.070**	−0.073***
	(0.028)	(0.028)	(0.028)
Pre-test score	−0.005	−0.002	−0.002
	(0.082)	(0.082)	(0.082)
Pre-test time	−0.004	−0.003	−0.004
	(0.023)	(0.023)	(0.023)
Pre-test trading app use_No, but	0.078**	0.071*	0.079**
	(0.036)	(0.037)	(0.036)
Pre-test trading app use_Yes	0.064	0.057	0.063
	(0.044)	(0.044)	(0.044)
Pre-test stated risk attitude	−0.002	−0.001	−0.002
	(0.008)	(0.008)	(0.008)
Pre-test SR	0.008	0.010	0.008
	(0.014)	(0.014)	(0.014)
Pre-test SA	−0.555***	−0.547***	−0.561***
	(0.066)	(0.066)	(0.066)
Female × T2	0.068		
	(0.058)		
WPI × T2		−0.001	
		(0.001)	
ISEE_Scholarship × T2			0.026
			(0.072)
ISEE_Mid-low × T2			−0.054
			(0.059)
Observations	217	217	217
R^2	0.353	0.353	0.354
Adjusted R^2	0.302	0.301	0.299
Residual Std. Error	0.185 (df = 200)	0.185 (df = 200)	0.186 (df = 199)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. Pre-test time and Pre-test SR enter in natural logarithms. Each column augments the fully controlled specification (Table 9, column 2) with pairwise interactions between the T2 indicator and a single moderator. Column headers indicate the moderator. Only treated students are included. Omitted ISEE category and abbreviations as in Table 18.

Table 21. Treatment effect heterogeneity: log post-test portfolio efficiency (consented subsample)

	Dependent variable:		
	Log post-test SR		
	(1)	(2)	(3)
	Gender	WPI	ISEE
T1	0.355**	−0.262	0.393*
	(0.166)	(0.608)	(0.235)
T2	0.412**	0.275	0.610**
	(0.167)	(0.604)	(0.246)
Female	−0.024	−0.206	−0.196
	(0.248)	(0.151)	(0.151)
Age	−0.009	−0.011	−0.007
	(0.033)	(0.033)	(0.033)
WPI	−0.001	−0.003	−0.001
	(0.002)	(0.004)	(0.002)
ISEE_Scholarship	−0.004	0.005	0.246
	(0.156)	(0.156)	(0.251)
ISEE_Mid-low	0.010	0.032	0.125
	(0.134)	(0.135)	(0.255)
Non-scientific HS	0.174	0.171	0.160
	(0.138)	(0.140)	(0.141)
High-dropout major	−0.170	−0.163	−0.167
	(0.123)	(0.124)	(0.123)
Pre-test score	0.382	0.395	0.419
	(0.376)	(0.376)	(0.376)
Pre-test time	0.024	0.033	0.034
	(0.099)	(0.098)	(0.098)
Pre-test trading app use_No, but	0.282*	0.269*	0.268*
	(0.155)	(0.158)	(0.156)
Pre-test trading app use_Yes	0.139	0.122	0.113
	(0.191)	(0.191)	(0.191)
Pre-test SR	0.466***	0.470***	0.456***
	(0.066)	(0.066)	(0.067)
Pre-test stated risk attitude	−0.065*	−0.060	−0.068*
	(0.037)	(0.037)	(0.037)
Pre-test SA	−0.290	−0.273	−0.301
	(0.302)	(0.304)	(0.303)
Female × T1	−0.323		
	(0.329)		
Female × T2	−0.189		
	(0.320)		
WPI × T1		0.005	
		(0.006)	
WPI × T2		0.001	
		(0.006)	
ISEE_Scholarship × T1			−0.415
			(0.368)
ISEE_Scholarship × T2			−0.410
			(0.369)
ISEE_Mid-low × T1			0.002
			(0.346)
ISEE_Mid-low × T2			−0.300
			(0.337)
Observations	319	319	319
R^2	0.230	0.229	0.234
Adjusted R^2	0.183	0.183	0.183
Residual Std. Error	1.011 (df = 300)	1.011 (df = 300)	1.011 (df = 298)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. The dependent variable is the natural logarithm of the post-test design-based Sharpe ratio. Pre-test time, and Pre-test SR enter in natural logarithms. Each column augments the fully controlled specification (Table 10, column 2) with pairwise interactions between treatment indicators and a single moderator. Column headers indicate the moderator. Omitted ISEE category and abbreviations as in Table 18.

Table 22. Treatment effect heterogeneity: log in-game portfolio efficiency (consented subsample)

	Dependent variable:		
	Log in-game SR		
	(1)	(2)	(3)
	Gender	WPI	ISEE
T2	0.379**	1.721**	0.764***
	(0.185)	(0.719)	(0.255)
Female	−0.366	−0.185	−0.179
	(0.280)	(0.205)	(0.205)
Age	0.009	0.006	0.007
	(0.042)	(0.042)	(0.042)
WPI	−0.002	0.004	−0.001
	(0.003)	(0.005)	(0.003)
ISEE_Scholarship	0.045	0.093	0.114
	(0.226)	(0.225)	(0.313)
ISEE_Mid-low	−0.019	0.026	0.323
	(0.181)	(0.182)	(0.259)
Non-scientific HS	−0.231	−0.285	−0.226
	(0.201)	(0.202)	(0.204)
High-dropout major	−0.308*	−0.246	−0.275
	(0.171)	(0.170)	(0.169)
Pre-test score	0.445	0.434	0.452
	(0.506)	(0.502)	(0.503)
Pre-test time	0.066	0.081	0.068
	(0.141)	(0.141)	(0.141)
Pre-test trading app use_No, but	0.282	0.214	0.302
	(0.222)	(0.223)	(0.221)
Pre-test trading app use_Yes	0.028	−0.026	0.018
	(0.272)	(0.271)	(0.270)
Pre-test SR	0.281***	0.293***	0.278***
	(0.088)	(0.088)	(0.087)
Pre-test stated risk attitude	−0.003	0.008	−0.008
	(0.050)	(0.050)	(0.050)
Pre-test SA	−0.372	−0.287	−0.394
	(0.403)	(0.404)	(0.402)
Female × T2	0.353		
	(0.354)		
WPI × T2		−0.012*	
		(0.007)	
ISEE_Scholarship × T2			−0.139
			(0.437)
ISEE_Mid-low × T2			−0.646*
			(0.358)
Observations	217	217	217
R^2	0.160	0.169	0.170
Adjusted R^2	0.093	0.102	0.100
Residual Std. Error	1.138 (df = 200)	1.132 (df = 200)	1.134 (df = 199)

Note: *p<0.1; **p<0.05; ***p<0.01. HC1-robust standard errors in parentheses. The dependent variable is the natural logarithm of the in-game design-based Sharpe ratio. Pre-test time, and Pre-test SR enter in natural logarithms. Each column augments the fully controlled specification (Table 11, column 2) with pairwise interactions between the T2 indicator and a single moderator. Column headers indicate the moderator. Only treated students are included. Omitted ISEE category and abbreviations as in Table 18.