# MindStock: Investigating How Principle-Anchored Feedback Supports Self-Reflection in Mobile Investment


Sooyohn Nam
Department of Industrial Design
KAIST
Daejeon, Republic of Korea
sooyohn.nam@kaist.ac.kr

Yeohyun Jung
Department of Industrial Design
KAIST
Daejeon, Republic of Korea
yeohyun@kaist.ac.kr

Kyuwon Cho
Department of Design
UNIST
Ulsan, Republic of Korea
unikyu20@unist.ac.kr

Hyunseung Lim
Department of Industrial Design
KAIST
Daejeon, Republic of Korea
charlie9807@kaist.ac.kr

Hwajung Hong
Department of Industrial Design
KAIST
Deajeon, Republic of Korea
hwajung@kaist.ac.kr


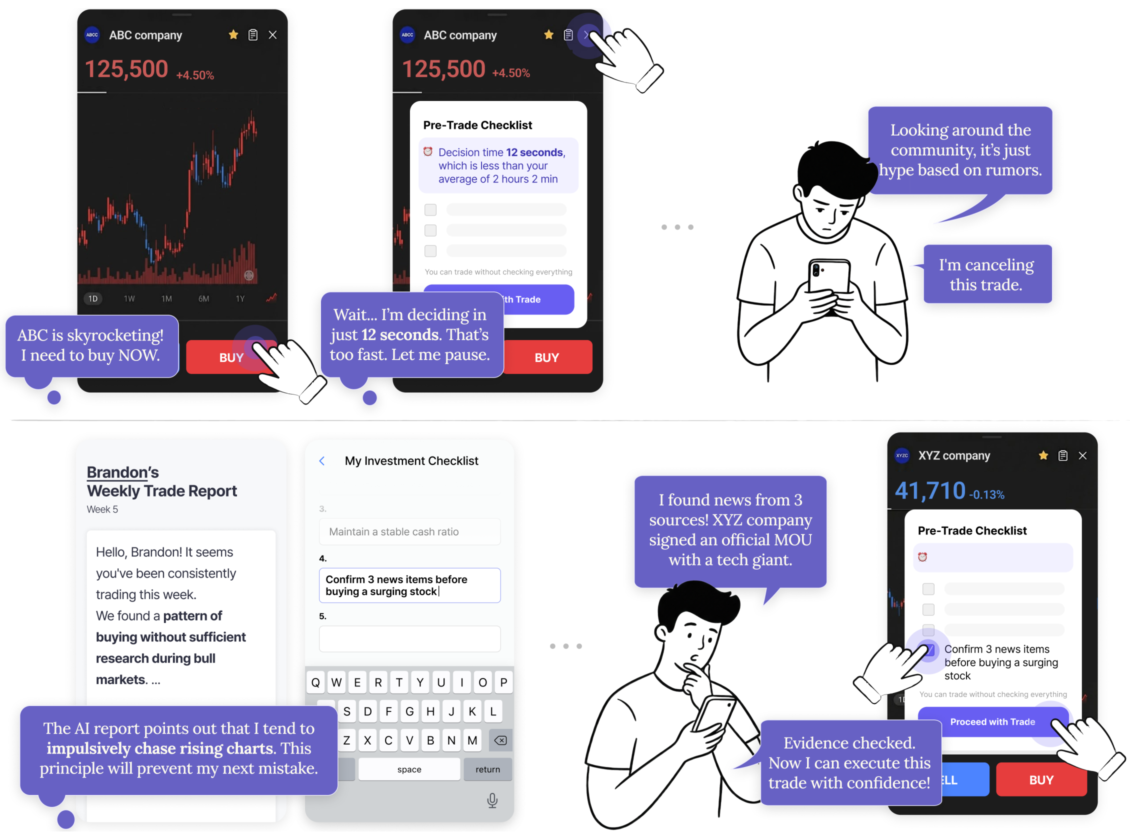


**Figure 1: We present MindStock, a technology probe providing principle-anchored feedback that combines user-defined investment principles with behavioral data mirroring trading patterns. The figure shows two scenarios: an investor canceling an impulsive trade upon recognizing unusually fast decision time (Top), and executing a researched trade after verifying alignment with self-defined principles informed by weekly pattern reports (Bottom).**

## Abstract


Investment decisions are often driven by time pressure and emotion, leaving investors vulnerable to cognitive biases. Mobile trading apps intensify these tendencies, yet existing interventions rely on external constraints that fail to foster lasting behavioral change. We investigate how reflection-centered approaches support mindful decision-making across a spectrum of investor expertise. We present




*DIS '26, Singapore, Singapore*



ACM ISBN 979-8-4007-2563-0/26/06
https://doi.org/10.1145/3800645.3812925

MindStock, a technology probe providing principle-anchored feedback by integrating user-defined principles with behavioral data mirroring trading patterns. In a 6-week field study with 16 investors, we found that meaningful reflection comes from the tension between principles and behavioral data. Principles give context to otherwise opaque metrics, while data keeps principles from drifting into vague self-assurances. This pattern varied by experience: novices gravitated toward normative rule-setting, while experienced investors used the system to test and refine their own assumptions. We contribute design implications for supporting reflection in high-stakes decision-making contexts.

## CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

## Keywords

Mobile Investment, Financial Decision Making, Mindful Investment, Data-driven Insight

**ACM Reference Format:**
Sooyohn Nam, Yeohyun Jung, Kyuwon Cho, Hyunseung Lim, and Hwajung Hong. 2026. MindStock: Investigating How Principle-Anchored Feedback Supports Self-Reflection in Mobile Investment. In *Designing Interactive Systems Conference (DIS '26), June 13–17, 2026, Singapore, Singapore.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3800645.3812925

## 1 Introduction

Investment is a high-stakes environment driven by time pressure and emotional arousal. Real-time price fluctuations compel investors to rely on heuristics rather than deliberate analysis [39, 50], leading to patterns such as loss aversion and overconfidence [3, 66]. Mobile trading apps intensify these tendencies by narrowing attention to top-gaining stocks [68] and enabling one-tap execution [10, 16], effectively prioritizing impulsive action over careful consideration.

To mitigate impulsive trading, HCI researchers have explored various interventions such as introducing friction [11, 16], warning messages [11]. For instance, Chaudhry and Kulkarni [11] proposed UI designs that introduce friction or warning messages to slow down trades. Singh et al. [61] investigated just-in-time interventions triggered by physiological signals, such as heart rate, to detect and mitigate emotional arousal. However, these external approaches may face challenges in investment contexts: outcomes are stochastic, and appropriate behavior varies across individuals depending on risk tolerance and strategy. This makes it difficult for systems to determine whether a given trade warrants intervention [5, 33].

These challenges suggest that supporting investors may benefit from approaches that help users recognize their own patterns rather than imposing external constraints. Reflection may be particularly valuable in this domain for several reasons. First, the probabilistic nature of investment outcomes makes it difficult to evaluate decision quality based on results alone [32, 66]. Nam et al. [47] observed that intermittent successes from impulsive trades can reinforce the very behaviors that warrant examination, creating cycles that are difficult to break through outcome observation. Second, as discussed above, external interventions risk triggering resistance when users perceive their autonomy as threatened. In response to these limitations, several researchers have emphasized the importance of designing technology that supports internal reflection rather than prescribing behavior [6, 54]. Baumer conceptualizes technology as a mirror that helps users see their own behavior more clearly, rather than as a tool that enforces normative judgments [6]. Yet while Personal Informatics research has demonstrated that data-driven self-reflection can support behavioral change [7, 37], applications to investment contexts remain relatively scarce. Examining the decision-making process itself may offer a path to interrupt maladaptive cycles, but there is limited empirical evidence on how such reflection-centered strategies operate in practice.

In this paper, we present *MindStock*, a mobile trading system designed to support investors' self-reflection through principle-anchored feedback. The system prompts users to articulate their own investment principles, which then serve as personalized baselines for interpreting behavioral data. MindStock implements this approach through: (1) an investment principles checklist that surfaces self-defined criteria at the point of trade execution, (2) pre-trade behavioral insights comparing current actions to personal baselines, and (3) weekly AI reports analyzing accumulated trading patterns.

We evaluate the system through a 6-week exploratory field study with 16 investors (8 novice, 8 experienced). Given that investment experience shapes decision-making tendencies and bias susceptibility [4, 18, 20], we also explore how this reflective mechanism operates across novice and experienced investors. Our findings suggest that reflection in this domain involves a dynamic negotiation between stated intentions and observed actions. We observe that principles without data tend to erode through self-rationalization, whereas data without principles can remain difficult to interpret. Meaningful reflection appears to emerge when principles provide the context for data, and data serves as evidence to verify compliance. We further find that this reflective mechanism plays out differently by experience: novices gravitated toward normative rule-setting, while experienced investors used the system to test and refine their own assumptions.

The major contributions of our study are as follows:

- The implementation and deployment of *MindStock*, a technology probe integrating reflection support into the mobile simulated trading workflow.
- Empirical evidence from a 6-week field deployment study examining how principle-anchored feedback influences investors' self-awareness and decision-making patterns, with distinct engagement patterns observed across novice and experienced investors.
- Design implications for financial systems to function as reflective partners, shifting focus from outcome metrics to process-oriented support.

## 2 Related Works

### 2.1 Design Patterns That Amplify Impulsive Trading

Mobile investment applications have expanded market access, enabling anyone to trade anytime and anywhere [10, 11]. However, many applications employ design patterns that prioritize transaction volume over decision quality. While some issues stem from mobile-specific constraints, others reflect broader design choices aimed at increasing user engagement and trading frequency. Mobile devices impose certain limitations. Wu and Wu [68] show that small screen sizes and limited information displays disproportionately concentrate users' attention on a narrow set of stocks, amplifying herding behavior as users chase highly visible top-gaining stocks. However, most problematic design patterns transcend platform boundaries. One-click execution removes deliberation time, collapsing the temporal gap between impulse and action [11]. Gamified elements such as confetti animations, achievement badges, and social leaderboards encourage frequent trading by treating investment as a competitive game rather than long-term wealth building [49]. In-app social feeds and continuous push notifications trigger emotional responses, intensifying fear of missing out (FOMO) or panic during market fluctuations [69]. These design choices, whether driven by business models that profit from transaction volume or by engagement optimization strategies, amplify well-documented behavioral biases including loss aversion, overconfidence, and impulsivity [11]. Prior work has shown that such design patterns systematically exploit users' cognitive biases to steer their behavior without their awareness [42], and that computing systems more broadly have the potential to induce and amplify biases in ways users often fail to recognize [8]. The result is systematic underperformance among individual investors [4, 48]. Researchers have increasingly criticized these applications for adopting design patterns similar to slot machines [49], calling for interventions that promote more deliberate investment behaviors [26, 27]. Responding to this concern, our research explores design interventions aimed at supporting reflective decision-making in trading contexts.

## 2.2 Existing Intervention Designs and their Potential Limitations

Researchers have proposed a range of interventions to support better financial decision-making. One line of work focuses on redesigning interface elements to mitigate bias-inducing features. For example, Chaudhry and Kulkarni [11] proposed design guidelines that remove slot machine-like visual elements and instead emphasize clear indicators of financial risk. Similarly, Kawai et al. [34] proposed integrating composite risk-duration metrics and educational warnings into leaderboard interfaces to counteract excessive short-term performance chasing. Another approach leverages persuasive games. For instance, Robinhood's Forest visualizes investment portfolios as growing forests, where impulsive short-term trading is represented as cutting down trees, while patient long-term investment allows the forest to flourish [12]. Likewise, Adventure Investor aims to reduce susceptibility to behavioral biases by embedding investment education within a game-based learning environment [56]. These existing interventions largely rely on external constraints (restricting certain actions or imposing delays) or explicit education (teaching users what constitutes good investment practice).

However, approaches that constrain user autonomy or prescribe specific behaviors face challenges in financial contexts. Lukoff et al. [40] demonstrate that tools which inhibit user agency can trigger psychological reactance, where individuals resist perceived external control to restore their sense of autonomy. In high-stakes domains like investment, users may view restrictions not as helpful guardrails but as impediments that prevent them from acting on time-sensitive opportunities, leading them to disengage from the system entirely [36]. Similarly, educational interventions that tell users what they should do may be perceived as patronizing or disconnected from individual circumstances and risk tolerance.

In response to these limitations, several researchers have emphasized supporting internal reflection rather than prescribing behavior [35, 41]. Reicherts et al. [54] argue that, particularly in high-stakes domains, cognitive support should help users think for themselves rather than making decisions on their behalf. Baumer [6] conceptualizes technology as a mirror that helps users see their own behavior more clearly, rather than as a tool that enforces normative judgments or rules. Building on this perspective, the present study aims to design an intervention that fosters investors' self-reflection beyond simple warnings or nudges.

## 2.3 Reflection-Centered Design and Its Applicability in Investment

The HCI community has developed methods to support self-reflection across domains such as health, productivity, and personal well-being. One common approach involves introducing intentional friction or visualizing personal data to prompt reflection on one's behavior [7, 19, 37]. For example, Cox et al. [16] concept of microboundaries disrupts unconscious habits by inserting small pauses before users take action, encouraging momentary awareness and deliberation. Visualizing personal data has also been shown to support deeper self-reflection. Choe et al. [14] found that exploratory visualization, which allows users to ask questions and discover patterns in their own data, leads to more meaningful insights than simple feedback alone. Thudt et al. [65] further demonstrated that constructive approaches to visualization, where users actively build representations of their data, can foster stronger engagement and reflection compared to passive consumption of pre-rendered charts. Another method focuses on Just-In-Time Adaptive Interventions (JITAIs), which aim to deliver support at moments when users are most receptive [46]. By intervening at critical moments of opportunity, JITAIs have been shown to significantly improve behavioral outcomes while minimizing disruption. Recent studies raised the need to extend these methods to financial contexts, proposing mindful friction to shift users from an impulsive hot state to a deliberative cool state [44, 47].

While these strategies are promising, the domain of financial investment presents distinct challenges. Financial markets are inherently stochastic environments in which the link between decision quality and outcomes is often severed [33]. A reckless, impulsive decision may yield profit due to luck, whereas a deliberate, reflective decision may still result in loss due to market volatility [5]. This complicates learning from experience and differentiates financial decision-making from other domains. As design research that integrates reflective interventions to promote better financial decision-making remains limited [11], there is a need to further investigate how reflection-centered interventions function in investment contexts. To this end, this study applies a reflection-centered intervention tailored to mobile investment contexts. Through this approach, we sought to investigate the impact on investors' recognition of emotional biases, understanding of trading patterns, and formation of long-term investment habits.

## 3 Design of MindStock

We designed MindStock as a technology probe to explore whether principle-anchored feedback can support investor self-reflection in mobile trading contexts. Following Hutchinson et al. [31]'s framework, we developed MindStock not as a finished product but as an instrument for understanding how investors engage with reflection-centered tools in naturalistic settings. Investment is inherently heterogeneous—investors differ in risk tolerance, strategy, and goals, and deliberate decision-making does not necessarily guarantee better outcomes [32]. In such a context, presenting behavioral data alone risks causing confusion or misinterpretation, as there is no universal correct behavior to measure against.

The system implements this approach through three core components: (1) a **Principles Checklist** that surfaces user-defined criteria at the point of trade execution, (2) **Pre-Trade Behavioral Insights** that provide objective behavioral metrics compared against personal baselines, and (3) a **Weekly AI Trade Report** that surfaces longitudinal patterns from accumulated trading data. Together, these components address different aspects of pre-trade awareness: the Principles Checklist reminds users of their self-defined investment principles at the moment of decision, while Pre-Trade Behavioral Insights surface objective behavioral data reflecting how their current actions compare to personal patterns. The Weekly AI Trade Report then enables delayed, pattern-based reflection outside the pressures of trading moments.

### 3.1 Principles Checklist

The Principles Checklist provides users an opportunity to check alignment with their self-defined investment principles just before execution. Prior work has suggested that investors prefer behavioral data over directive interventions and value both real-time support and retrospective analysis capabilities [47]. Building on this insight, rather than imposing external standards, the system prompts users to articulate their own investment principles at the outset. These self-defined principles then serve as a personalized reference frame through which behavioral data can be interpreted.

This component draws on the concept of implementation intentions, specific if-then plans that link situational cues to intended responses [21]. By prompting users to articulate concrete criteria during onboarding and surfacing them at the decision point, the checklist aims to bridge the gap between intentions formed in calm states and decisions made under pressure. Users can proceed with trading without checking all items, preserving user autonomy while enabling observation of individual usage patterns. This design allows the system to log which items users skip, how checking behaviors differ across participants, and what the average compliance rate is.

During onboarding, users author 3-5 personal investment rules (Figure 3A). The system provides example principles for reference (e.g., I will not buy stocks I haven't researched for at least 3 days, I will set a stop-loss before buying). Users can freely modify, delete, or add these principles in the Settings tab. This authoring process serves as an initial reflection opportunity, prompting users to externalize criteria that may otherwise remain implicit.

The checklist appears once when users press the execution button on the order confirmation screen (Figure 3B). This timing represents the final decision point where users have made their buy/sell decision but have not yet executed it. Prior research suggests that moments just before trade execution may offer opportunities for reassessment [47]. Since a system-structured pause (the order confirmation procedure) already exists at this moment, a reflection opportunity can be inserted while minimizing additional friction [16].

Users can proceed with trading without checking all items. This design preserves user autonomy while enabling observation of individual usage patterns—which items users skip, how checking criteria differ across participants, and what the average compliance rate is. Each checklist interaction—checked items, skipped items, and viewing time—is logged for analysis. Examples of participant-authored checklists are provided in Appendix A.

### 3.2 Pre-Trade Behavioral Insights

While principles provide a self-defined framework for decision-making, objective behavioral data can reveal discrepancies between intentions and actions. However, behavioral finance research demonstrates that biometric or behavioral signals require interpretation within individual context [3]. Pre-Trade Behavioral Insights addresses this by providing personalized baselines against which current behaviors can be compared.

The system tracks five behavioral metrics selected based on two criteria: (1) signals that can indicate cognitive biases frequently occurring in investment [32, 66], and (2) information computable in real-time using only action logs and smartphone sensors.

**Decision Time** measures the total wall-clock time elapsed from the user's first access to a specific stock's detail page to the final execution of the trade, explicitly including all offline periods. We operationalized this metric to capture overall behavioral pace rather than precise cognitive duration. Drawing on behavioral decision theory, abnormally short decision times suggest reliance on fast, impulsive thinking instead of deliberative thought processes, while excessively long times may indicate hesitation or analysis paralysis [32]. We adopted this wall-clock measurement as an indicator to infer potential impulsivity.

**Trade Size** compares the currently intended trade volume with the user's typical position size. Deviations may indicate cognitive biases affecting risk perception—abnormally large positions may signal overconfidence, whereas abnormally small positions can reflect loss aversion stemming from a desire to avoid regret [66].

**Trading Frequency** compares the number of trades executed during the current session or day with the user's historical average. Behavioral finance research indicates that high trading frequency is often driven by emotional reactions to short-term market noise and overconfidence rather than fundamental information, leading to overtrading [3, 48].

**Hand Tremor** compares the standard deviation of accelerometer readings measured at the order confirmation screen with the user's baseline. This metric serves as a potential physiological proxy for acute arousal, stress, and anxiety that can be measured unobtrusively through smartphone sensors. HCI research suggests that psychological stress can induce muscle tension, which may manifest as subtle tremors or irregular movements during mobile interactions [29, 57, 63], especially under conditions of negative emotional arousal such as frustration [51].

**News Checking** compares the number of news articles viewed within the MindStock app before a trade with the user's usual pre-trade research pattern. Deviations show how limited attention resources are allocated under pressure—abnormally high checking may indicate anxiety-driven information seeking, while abnormally low frequency can signal impulsive decisions made without proper due diligence [30].

This intervention was implemented in two phases of the 6-week study. In Phase 1 (weeks 1-3), no behavioral insights were shown, establishing personal baselines. Starting in Phase 2 (weeks 4-6), users could select 2-3 out of 5 metrics in app settings to display just before trading.

Just before trading, the selected metrics appear alongside the Principles Checklist on the order confirmation screen (Figure 3C). Each metric employs statistical thresholding: values within ±1 standard deviation of the user's personal mean are described as "similar to your average," while values beyond ±1 standard deviation are labeled as "less than" or "greater than" the average. Importantly, no value judgments are attached to these labels—"less than average" is not framed as inherently good or bad, leaving interpretation to the user.

## 3.3 Weekly AI Trade Report

Real-time interventions alone may be insufficient for meaningful reflection. Research on self-regulated learning suggests that stepping back from immediate task demands enables deeper understanding of one's own behavioral patterns [70]. Moreover, the pressure and emotional arousal inherent in trading moments make in-the-moment reflection difficult [39].

To capture how emotions evolve around trading decisions, the system prompts users at two points in time. Research on reflection suggests that revisiting past experiences from a temporal distance can reveal patterns that are difficult to notice in the moment [19]. Immediately after each trade, users select their emotions and briefly describe what drove their decision (Figure 3D-1). Then, 72 hours later, a follow-up prompt asks how they feel about the same trade now (Figure 3D-2). By comparing these two responses, the weekly report can surface patterns in how emotions shift after trading, such as recurring regret or overconfidence, that users often miss in the moment.

The Weekly AI Trade Report provides non-prescriptive, AI-driven analysis of accumulated trading patterns (Figure 3E). Every Monday at 9:00 AM, users receive a web link to their personalized report via KakaoTalk. The report analyzes the past week's trading data, allowing users to step back from temporal and emotional pressure and understand their behavior from a macro perspective in a calmer state. Figure 2 illustrates the end-to-end pipeline for generating the Weekly AI Trade Report.

To provide a macro view of users' investment behavior, we employ itemset mining. This method surfaces behaviors that frequently co-occur without imposing temporal order. For instance, the system might identify: "When buying stocks, you often: view detailed information + check charts + skip checklist + execute quickly." We selected itemset mining because in investment decision-making, which information sources are consulted together matters more than the order in which they are accessed [50].

Each report follows a structured format:

- **Pattern Summary**: A brief overview of the most prominent behavioral pattern discovered with frequency.
- **Detailed Pattern Analysis**: Breakdown of co-occurring behaviors, including how checklist usage, decision speed, emotional states (from immediate and 72-hour post-trade reflections), and market context relate to the pattern.
- **Self-Reflection Questions**: Open-ended prompts encouraging users to explain their reasoning (e.g., "What might have caused you to skip the checklist when this pattern occurred?" or "How do your emotions immediately after trading compare to how you feel three days later?").
- **Personalized Checklist Suggestions**: Actionable checklist items tailored to improve observed behavioral patterns, connecting with users' existing principles where applicable.

The AI does not prescribe actions or offer investment advice. Instead, it describes observed patterns and poses questions. Self-explanation questions are grounded in educational research showing that when learners articulate reasoning in their own words, they develop deeper understanding and better self-regulation [13]. The complete system prompt is provided in Appendix D.

The weekly cadence represents a deliberate choice to provide delayed feedback complementary to real-time insights. Educational research indicates that while immediate feedback corrects errors in the moment, delayed feedback can enhance long-term retention and understanding by requiring users to recall and reconstruct situations [60]. The weekly interval provides emotional distance, enabling more objective analysis of trading behaviors.

## 3.4 Implementation

We implemented MindStock as a mobile trading application using React Native [43] with TypeScript [45]. We embedded AlphaSquare—a commercial Korean stock information platform—within the app using react-native-webview [52]. AlphaSquare provides market data and a paper trading feature that allows users to execute simulated trades with virtual currency. The application uses Firebase [22] for authentication, data storage (Firestore [23]), and serverless computation (Cloud Functions [24]).

The app collects six categories of behavioral data through event logging: (1) trading flow—timestamps from initial stock viewing to trade confirmation or cancellation; (2) screen dwell time—duration spent on different views; (3) information exploration—interactions with news articles and reading duration; (4) sensor data—hand tremor via accelerometer using react-native-sensors [53]; (5) checklist usage; and (6) retrospective reflection—emotional states captured immediately and 72 hours post-trade. We also record market context including stock prices and index movements at trade time.

For weekly pattern discovery, we employ frequency-based itemset mining [1, 64] to identify co-occurring behaviors within trading episodes. A scheduled Cloud Run [25] job performs itemset mining

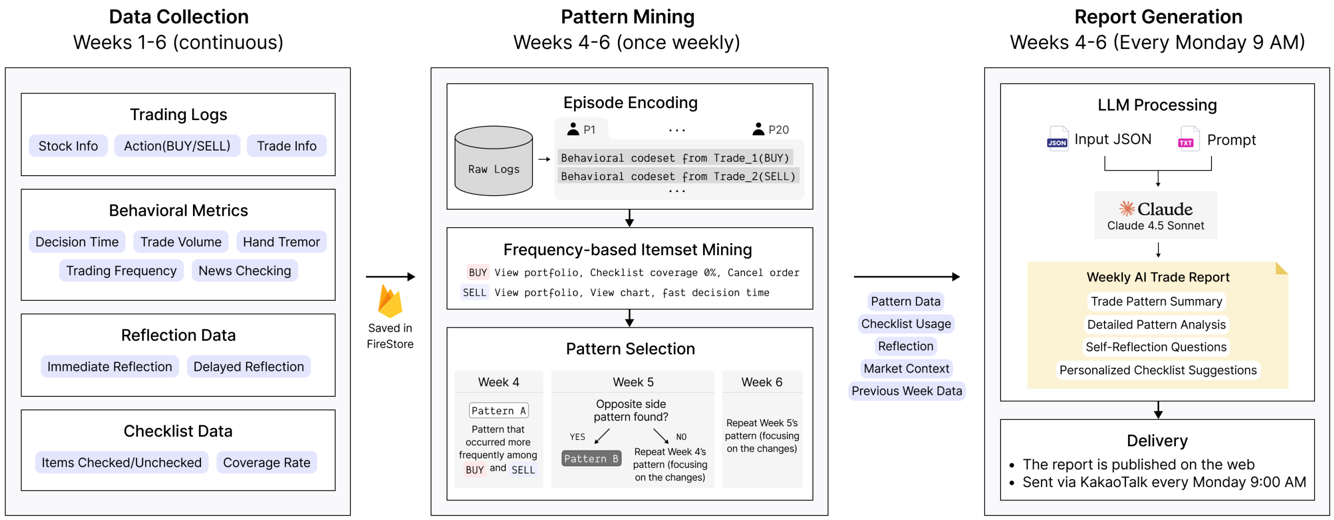


**Figure 2: Weekly AI Trade Report generation pipeline. Data is continuously collected during Weeks 1–6 and stored in Firestore. Starting in Week 4, a weekly pattern mining process encodes each user's trades into behavioral episodes, identifies the highest-frequency patterns for BUY and SELL actions separately, and selects patterns according to a protocol that aims to cover both trade types. The selected patterns, along with checklist usage, reflection data, and market context, are formatted as input JSON and processed by Claude 4.5 Sonnet to generate personalized reports delivered via KakaoTalk every Monday at 9:00 AM.**

separately for buy and sell trades for each user. Selected patterns are formatted as structured JSON and sent to Claude 4.5 Sonnet via the Anthropic API [2] with a system prompt that instructs the model to describe behavioral patterns in observational language, interpret deviations from personal baselines rather than raw values, and pose open-ended reflection questions rather than prescribing actions. Generated reports follow a five-section structure in Korean: pattern summary, behavioral analysis, emotional characteristics, self-reflection questions, and personalized checklist suggestions. The complete system prompt is provided in Appendix D.

Pre-Trade Insights calculates personal baselines from users' trading history. At the order confirmation screen, current metrics are compared to baselines using ±1 standard deviation thresholds, with deviations described neutrally.

## 4 Field Study

To evaluate how MindStock's principle-anchored reflection approach influences investors' decision-making patterns and self-awareness, we conducted a 6-week field study. Rather than comparing MindStock against alternative interventions, we adopted an exploratory deployment to observe how investors engage with reflection-centered tools in realistic trading contexts.

We employed simulated trading because our research questions center on how investors engage with reflection tools, not on financial outcomes per se. Since MindStock introduces a mandatory cognitive pause into the trading flow, we needed an environment where participants could fully experience this intervention without the pressure of real financial consequences suppressing exploratory behavior. Real stakes would likely cause participants to either abandon the reflection process to minimize latency or avoid risky trades altogether, both of which would obscure the behavioral patterns we aimed to study. Additionally, deploying an untested intervention with real capital raises ethical concerns, as intervention-induced delays could contribute to actual financial losses. Using controlled or simulated environments to study risky decision-making while ensuring participant safety is an approach adopted by prior studies examining financial and complex decision-making behaviors [12, 54]. To maximize ecological validity within this simulated context, we followed Dole and Ju's immersive simulation guidelines [17], which suggest that experimental realism may serve as a valid proxy when in-situ measurement is infeasible. We embedded a commercial platform (AlphaSquare) identical to real trading interfaces and measured participants' sense of presence using a standardized questionnaire. The study protocol was approved by our institution's Institutional Review Board (IRB).

### 4.1 Participants

We recruited participants through university online communities and investment forums, supplemented by snowball sampling. Eligible participants were Korean adults (age ≥20) who owned iOS devices (≥15.1), were active investors capable of trading daily, and were available for the entire 6-week study period. A total of 20 participants enrolled; however, 4 (P10, P11, P16, P17) were excluded from analysis due to non-compliance (no login for ≥4 consecutive weekdays or total trades <15). The final sample consisted of 16 participants ($M$ = 25.8 years, $SD$ = 2.1; 7 females, 7 males, 2 undisclosed) with an average investing experience of 31.5 months ($SD$ = 19.4).

Behavioral finance research suggests that investment experience shapes both decision-making mechanisms and the types of biases

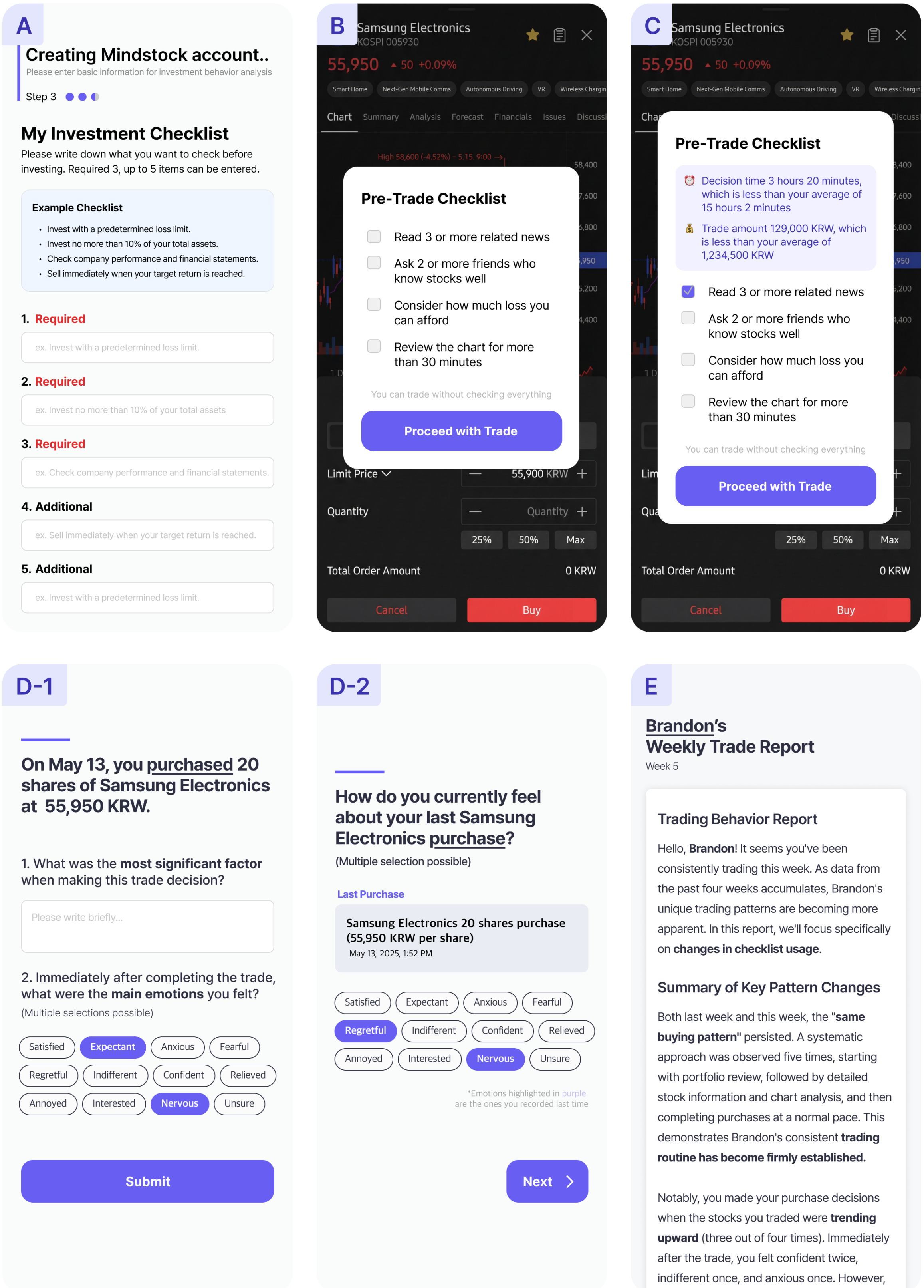


**Figure 3: Overview of MindStock. (A) Principles checklist initialization. (B) Principles checklist popup (Phase 1). (C) Principles checklist with Pre-Trade Behavioral Insights (Phase 2). (D-1) Immediate post-trade reflection. (D-2) Delayed reflection (72 hours later). (E) Weekly AI Trade Report.**

to which investors are vulnerable [18]. Novice investors, lacking structured decision criteria, tend to be susceptible to emotional reactivity and herd behavior [4, 38], whereas experienced investors' accumulated knowledge can paradoxically lead to overconfidence and confirmation bias [20]. By distinguishing between these two groups with distinct vulnerability profiles, we aimed to examine how principle-anchored reflection operates differently across user types.

To operationalize these experience differences, we developed a 4-item screening checklist (Table 1). Each item was validated through consultations with financial UX experts from the UX design team of a major investment and securities firm. Participants meeting ≥3 criteria were classified as Experienced ($n$ = 8); others as Novice ($n$ = 8). We adopted multiple criteria rather than a single threshold because investment expertise is not defined by duration alone, but encompasses multiple dimensions including trading frequency, responsiveness to market conditions, and portfolio management practices.

Of the 20 initially enrolled participants, four were excluded from analysis. Two were excluded due to insufficient in-app trading activity, rendering them unsuitable for testing a reflection system that relies on active decision-making logs; the other two failed to meet minimum engagement requirements during the study period. Each participant received 120,000 KRW (approximately $85 USD) as compensation. Detailed demographics are provided in Table 2.

## 4.2 Procedure

The study consisted of three stages: pre-survey and orientation, a 6-week field deployment, and exit survey with interview (Figure 4). The 6-week deployment was divided into two phases with different intervention components activated in each phase (Table 3).

***Pre-survey and Orientation.*** Before orientation, participants completed an online pre-survey assessing investment metacognitive awareness using an 18-item questionnaire adapted from the Metacognitive Awareness Inventory [58]. This measure was included to track changes in participants' self-reported metacognition over the study period, enabling examination of whether engagement with principle-anchored reflection corresponds to shifts in self-awareness. The scale measures two dimensions: knowledge of cognition (e.g., "I am aware of situations where I make impulsive buy decisions") and regulation of cognition (e.g., "I have clear criteria for when to sell"). All items used 7-point Likert scales. The complete questionnaire items are provided in Appendix B.

Orientation sessions were conducted via Zoom in groups of 3-5 participants, lasting exactly 30 minutes. Researchers explained the study procedure, assisted with TestFlight installation, and demonstrated key features: principles checklist customization, behavioral insights (explaining the five metrics and their calculation methods, accessible in app settings), weekly AI reports, and post-trade reflection prompts. Participants were informed that Phase 1 (weeks 1-3) would show only the checklist, while Phase 2 (weeks 4-6) would activate behavioral insights and weekly reports. Users were instructed that they could select 2-3 of the 5 behavioral insight metrics to display before each trade, though all five would remain accessible in settings and would appear in post-trade summaries during Phase 2.

***6-week Field Deployment.*** The phased intervention structure was designed to systematically examine the mechanisms of principle-anchored reflection by progressively introducing components that differ in temporal scale and cognitive function. The Principles Checklist supports momentary intention alignment before execution, Pre-Trade Behavioral Insights provide real-time signals indicating the user's current state, and Weekly AI Trade Reports facilitate delayed pattern-based reflection outside the pressures of trading moments. Providing all components simultaneously would make it difficult to disentangle how each contributes to the reflective process.

Phase 1 (weeks 1–3) deployed only the Principles Checklist, serving two purposes: first, allowing examination of how users construct reflection based solely on self-authored principles and subjective judgment; second, establishing personalized behavioral baselines necessary for the comparative metrics in Phase 2. This baseline period was essential for Pre-Trade Behavioral Insights to compute individualized thresholds (e.g., decision time, trade size relative to personal averages). Phase 2 (weeks 4–6) introduced Pre-Trade Behavioral Insights and Weekly AI Trade Reports in the same trading context, enabling observation of how the reflective process expands when subjective checks are supplemented with behavioral state information and accumulated patterns.

After completing each trade, participants encountered a summary screen displaying transaction details and reflection prompts. In Phase 1, participants viewed trade information (stock name, quantity, price) and responded to two prompts: a free-text question asking what factors most influenced their decision, and an emotion selection from 12 options (e.g., expectation, anxiety, confidence, regret—multiple selections allowed). In Phase 2, participants additionally viewed all five behavioral insights on this summary screen, regardless of which metrics they had selected for pre-trade display.

Seventy-two hours after each trade, participants encountered a delayed reflection prompt upon their next app login, asking how they currently felt about their recent purchase. This in-app popup presented the same 12 emotion options, enabling participants to compare their immediate and delayed emotional responses.

During weeks 4–6, participants received weekly AI-generated reports analyzing their trading patterns. Each Monday at 9:00 AM, participants received a KakaoTalk message containing a web link to their personalized report. After reviewing the report, participants could optionally respond to feedback questions assessing whether the identified patterns matched their actual behavior and whether they found the report useful.

Throughout the study, researchers sent weekly encouragement messages every Monday via group chat on KakaoTalk, but otherwise maintained hands-off observation except for technical support. Participants could not disable intervention features, though they could skip individual checklist prompts.

Mid-study assessment occurred at the end of week 3, repeating the 18-item metacognitive awareness questionnaire to capture changes at the phase transition.

***Exit Survey and Interview.*** After completing week 6, participants completed a post-survey including: (1) the metacognitive awareness questionnaire (third administration), (2) the Presence Questionnaire based on the UQO Cyberpsychology Lab's revised version [67],

| Criterion | Rationale | Ref. |
|---|---|---|
| ≥2 years of actual stock investment experience | Longer market exposure allows investors to accumulate experiential knowledge, internalize feedback from past outcomes, and develop more stable decision-making patterns—characteristics commonly associated with experienced investors. | [59] |
| ≥10 sell decisions in the past year | Executing sell decisions requires deliberate evaluation of gains, losses, and future expectations; investors who perform sell decisions frequently demonstrate more active portfolio management and greater engagement with the full investment cycle. | [48] |
| Continued investing during market downturns | Investors who remain active during downturns tend to possess greater risk tolerance, market familiarity, and resilience to volatility—traits distinguishing experienced investors from those who withdraw under adverse conditions. | [28] |
| Currently holding ≥3 different stocks | Maintaining multiple positions indicates awareness of diversification and broader portfolio management strategies, reflecting more sophisticated investment practices typically observed among experienced investors. | [62] |

**Table 1: Experience classification criteria. Thresholds were operationalized through practitioner consultation, based on empirically-supported investor characteristics.**

| ID | Gender | Age | Occupation | Investing Experience | Group |
|---|---|---|---|---|---|
| P1 | F | 25 | Office Worker | 13 months | Novice |
| P2 | M | 26 | Master's Student (Industrial Design) | 3 years | Experienced |
| P3 | F | 26 | Master's Student (Computer Science) | 3 years | Novice |
| P4 | M | 29 | Ph.D Student (Industrial Design) | 5 years | Experienced |
| P5 | M | 26 | Graduate Student | 3 years | Experienced |
| P6 | M | 26 | Ph.D Student (Computer Science) | 2 months | Novice |
| P7 | M | 30 | Master's Student (Chemical & Biomolecular Engineering) | 2 years | Novice |
| P8 | – | 25 | Freelancer | 5 years | Experienced |
| P9 | M | 27 | Master's Student (Industrial Design) | 2 months | Novice |
| P12 | M | 23 | Undergraduate (Chemical & Biomolecular Engineering) | 2 years | Novice |
| P13 | F | 27 | Software Developer | 3 years | Experienced |
| P14 | – | 24 | Undergraduate (Computer Science) | 5 years | Novice |
| P15 | F | 28 | Undergraduate (Pre-Medicine) | 7 months | Novice |
| P18 | F | 24 | Aerospace Researcher | 3 years | Experienced |
| P19 | F | 24 | Ph.D Student (Economics) | 4 years | Experienced |
| P20 | F | 22 | Designer | 2 years | Experienced |

**Table 2: Participants' demographics and investing experience. P10, P11, P16, and P17 were excluded due to non-compliance (no login ≥4 days or total trades <15).**

adapted for investment contexts to assess the perceived realism of the paper trading environment. The complete adapted Presence Questionnaire items are provided in Appendix C.

Individual exit interviews were conducted remotely via Zoom, lasting approximately 60 minutes each. Interviews followed a structured protocol covering: (1) overall study experience, (2) behavior-specific questions derived from each participant's log data (e.g., "You rarely checked [specific checklist item]—why was this principle difficult to follow?"), (3) perceived effectiveness of each intervention component, and (4) likelihood of transfer to real investment behavior. All interviews were audio-recorded with consent and transcribed verbatim for analysis.

## 4.3 Data Analysis

Given the exploratory nature of this deployment without control conditions, we prioritized qualitative analysis to understand how investors engaged with principle-anchored reflection. Quantitative comparisons between Phase 1 and Phase 2 risk spurious conclusions due to uncontrolled confounds including learning effects, market volatility, and temporal factors. Therefore, we treat quantitative data primarily as descriptive measures to contextualize qualitative findings. While we avoid making causal claims about behavioral outcomes (e.g., profitability) due to external confounds, we utilize non-parametric tests on self-reported measures to identify significant shifts in participants' metacognitive awareness over the study period.

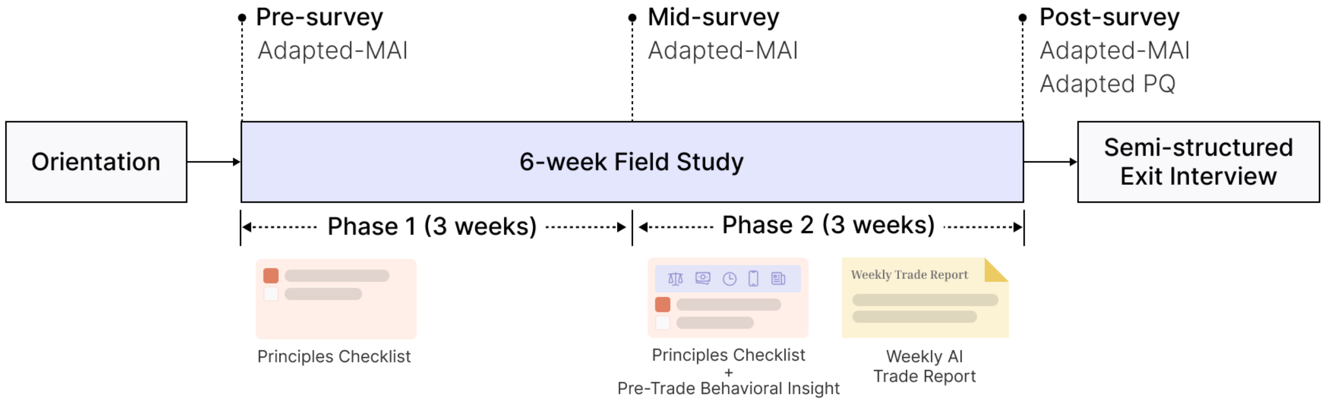


**Figure 4: Field study timeline. The 6-week deployment was divided into two phases: Phase 1 (weeks 1-3) with Principles Checklist only, and Phase 2 (weeks 4-6) with full intervention components including Pre-Trade Behavioral Insights and Weekly AI Trade Reports. Pre-survey, mid-study assessment, and exit survey/interview bookended the deployment.**

| Intervention Component | Phase 1 (Weeks 1-3) | Phase 2 (Weeks 4-6) |
|---|---|---|
| **Principles Checklist** | Yes | Yes |
| **Pre-Trade Behavioral Insights** | No | Yes |
| **Post-Trade Reflection Prompts** | | |
| - Immediate reflection (trade details only) | Yes | Yes |
| - Immediate reflection (with Behavioral Insights) | No | Yes |
| - Delayed reflection (72 hours later) | Yes | Yes |
| **Weekly AI Trade Report** | No | Yes |
| **Background Data Collection** | Yes | Yes |

**Table 3: Intervention components activated in each phase of the 6-week field study.**

***Application Log Data.*** Behavioral metrics were extracted from Firestore logs to characterize participant engagement. We report descriptive statistics for key indicators including: trading frequency, checklist compliance rate, reflection completion rate (immediate and delayed), and behavioral insight metric selection. These metrics serve to contextualize qualitative findings rather than to evaluate intervention effects.

***Survey Data.*** The 18-item metacognitive awareness questionnaire was administered at three time points (pre, mid, post) to track self-reported changes in investment metacognition. To analyze differences across the three time points, we used the Friedman test, a non-parametric alternative appropriate for repeated measures with ordinal data. For items showing significant differences, we conducted post-hoc pairwise comparisons using Wilcoxon signed-rank tests with Bonferroni correction. The Presence Questionnaire was analyzed using descriptive statistics to assess the perceived realism of the simulated environment. We report these survey results descriptively to complement interview findings.

***Interview Data.*** Interview transcripts served as the primary data source and were analyzed using thematic analysis [9]. The first author conducted open coding in Atlas.ti, identifying initial concepts related to metacognitive changes, intervention component experiences, and behavioral adaptations. Codes were then grouped into higher-level categories through iterative refinement. One additional researcher reviewed the coding and provided feedback, leading to iterative revision by the entire research team.

# 5 Findings

## 5.1 Engagement and Metacognitive Changes

Across the 6-week study period, 16 participants executed 703 trades, generating a rich log of behavioral changes. Daily trading activity reached a peak in the first week as users explored the system, followed by stabilization from weeks two to five, as illustrated in Figure 6. A five-day rolling average shows the transition from initial high engagement to a more consistent trading routine. Trading activity varied by experience group, as novice traders executed 440 trades and experienced traders executed 263. Reflection completion rates remained high across both groups, with novice traders completing 70.7% and experienced traders completing 60.1% for immediate reflection. This indicated sustained engagement with the intervention despite the friction it introduced (Table 4).

Participants reported reasonable engagement with the simulated trading environment. The Presence Questionnaire results showed a mean of 99.31 with a standard deviation of 9.52, while the normative reference reached $M$ = 104.39 and $SD$ = 18.99. While overall presence scores were slightly below the norm, the Realism subscale reached $M$ = 37.06 ($SD$ = 4.75) and exceeded the normative reference of $M$ = 29.45 ($SD$ = 12.04). This suggests that participants perceived the trading interactions and market dynamics as believable despite the simulated stakes.

We compared participant responses on the 18-item investment metacognition questionnaire across three time points including Pre, Mid, and Post using the Friedman test. The results indicated statistically significant differences in 11 items ($p < .05$). To identify which time points differed significantly, we conducted post-hoc Wilcoxon signed-rank tests with Bonferroni correction ($\alpha = .0167$). Significant improvements were observed particularly in areas related to self-awareness and principle refinement. For instance, scores for item Q3, which assesses the ability to name investment habits, significantly increased from a Pre mean of 3.69 ($SD$ = 1.45) to a Post mean of 5.50 ($SD$ = 0.61, $p < .001$). Similarly, scores for item Q18 on the refinement of principles rose from 3.69 to 5.31 ($p < .001$). Items Q11, Q12, Q16, and Q17 regarding the establishment of and adherence to principles also showed significant improvements in the post-intervention phase compared to the baseline.

## 5.2 Checklists as Cognitive Interruptions and Negotiable Rules

*5.2.1 Establishing Principles Grounded in Experience.* Users treated the checklist as a canvas for externalizing tacit criteria they had never written down before. While 44% of items followed the provided examples, 56% were newly created, often using personal language to concretize psychological goals. P2, for instance, set a rule to *"not trust my gut too much,"* grounding an abstract self-awareness into a concrete trading criterion.

The resulting checklists fell into three functional categories. **Rule Setting** (38%) established behavioral constraints, such as P8's rule to *"set a loss limit before investing."* **Emotional Regulation** (38%) targeted internal states, including P4's *"do not feel FOMO for rising stocks."* **Evidence Seeking** (24%) focused on verifying justifications, such as P9's rule to *"think of at least three investment rationales."* This distribution suggests that users prioritized controlling behavior and managing internal states, focusing primarily on the emotional and self-regulatory challenges that dominate investment decision-making.

*5.2.2 Internalizing Principles through Deliberate Action.* Checking items on the list was just as important as the rules themselves because it forced a moment to think that most trading apps do not provide. Although this interruption felt uncomfortable at first , participants eventually viewed it as a helpful space to stop and think. For example, P2 mentioned that the extra step created a buffer to consider whether buying a stock was the right choice at that moment. This was especially useful for those who recognized their own impulsive tendencies. P9, who often bought stocks on impulse, found that the popup helped them calm down and reconsider their actions. Several participants reported that the checklist directly led to different decisions and helped them build a habit of thinking once more before executing trades.

Using the checklist repeatedly changed it from a rule set by the system into a personal promise. Even when behavior did not change right away, the criteria remained in the minds of the participants and influenced how they approached future trades. P2 described this effect as planting seeds in the mind that might grow over time. The process of writing the principles also served as an early reflection. P4 noted that authoring the checklist made the rules stay in their head for the entire six weeks. This suggests the checklist functioned as a summary of personal patterns, which is a role distinct from a basic list of rules.

*5.2.3 Self-Rationalization and Rule Loosening.* Sustained use also revealed a tension where the more familiar the checklist became, the easier it was to check items without really thinking about them. Despite these benefits, quantitative logs and qualitative interviews revealed patterns of self-rationalization. Over time, strict compliance often gave way to loose interpretation. This pattern is corroborated by the quantitative logs shown in Figure 5 (Left). While novice traders began with high compliance rates of 83.9% in Phase 1, this dropped significantly to 68.7% in Phase 2. P12 admitted, *"I think I've just checked everything and made the trade... rather than trying to follow it... I just checked without thinking."* This illustrates how checking became a formality.

Criteria also became vague over time. Participants began satisfying checklist items based on fragmentary impressions rather than concrete evidence, checking off rules they had only loosely followed in their heads. P13 illustrated this pattern clearly, checking off their loss limit item when they *"roughly thought in my head 'I should cut losses on this at about this point'"* instead of adhering to a specific threshold. These patterns suggest that self-authored rules, when relying solely on subjective judgment, can gradually drift toward loose interpretation over time.

## 5.3 Behavioral Data as Both Mirror and Enigma

*5.3.1 Real-Time Signals of the Hot State.* The Pre-Trade Behavioral Insights functioned as a mirror, visualizing subtle behaviors users ignored in the "hot state" of trading. Metrics such as decision time and trade volume provided concrete values for moments participants felt they "just did." When connected to principles, these metrics effectively bridged the gap between intention and action. P19 described how decision time data prompted reconsideration: *"There were times when I was about to sell... but when I saw that the decision time seemed a bit fast, I went back, checked the chart once more, and then sold."* Similarly, P14 used the trade volume metric to monitor adherence to their installment buying principle and found it *"helpful"* to know their standing relative to their average.

Other metrics revealed unexpected aspects of users' behavior. P18 reflected only after seeing the data: *"I didn't know I had hesitated... but seeing that it took a bit longer than usual... I reflected like 'I hesitated a bit. Was I impulsive, did I lack confidence?'"* Hand tremor data served as a particularly powerful signal of physical tension. P5 was surprised to confirm that *"my hands actually trembled a bit... when I really make challenging trades."* This countered

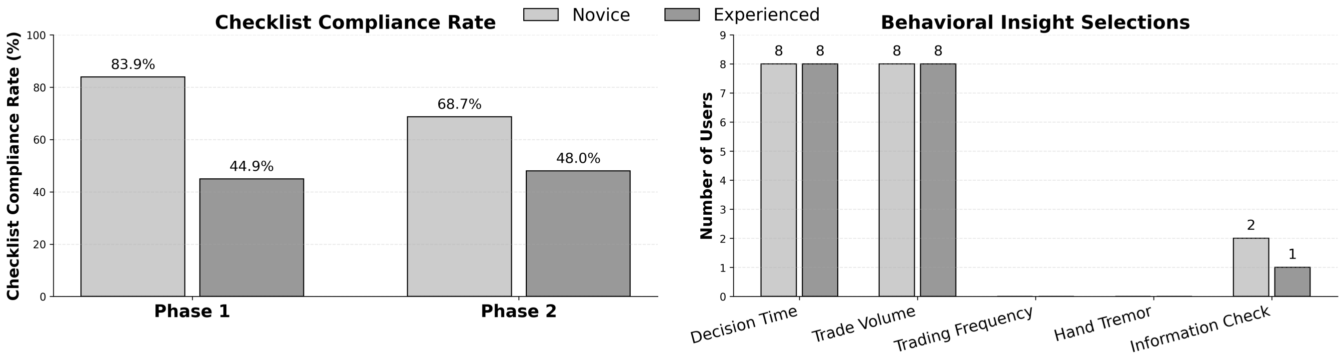


**Figure 5: Checklist compliance rates and behavioral insight selections by experience group. Left: Checklist compliance rates across phases. Novice traders showed higher initial compliance (83.9%) that decreased in Phase 2 (68.7%), while experienced traders maintained relatively stable rates (44.9% to 48.0%). Right: Number of participants who selected each behavioral insight for pre-trade display during Phase 2. All 16 participants retained the two default metrics (Decision Time and Trade Volume), while no participant selected Trading Frequency or Hand Tremor. Only 3 participants (2 novice, 1 experienced) added Information Check.**

| Group | Users | Trades | Immediate | Delayed | Delayed \| Immediate |
|---|---|---|---|---|---|
| Novice | 8 | 440 | 311 (70.7%) | 251 (57.0%) | 251/311 (80.7%) |
| Experienced | 8 | 263 | 158 (60.1%) | 139 (52.9%) | 139/158 (88.0%) |
| **Overall** | 16 | 703 | 469 (66.7%) | 390 (55.5%) | 390/469 (83.2%) |

**Table 4: Reflection completion rates by experience group. Delayed reflection was prompted 72 hours later only for trades where immediate reflection was completed.**

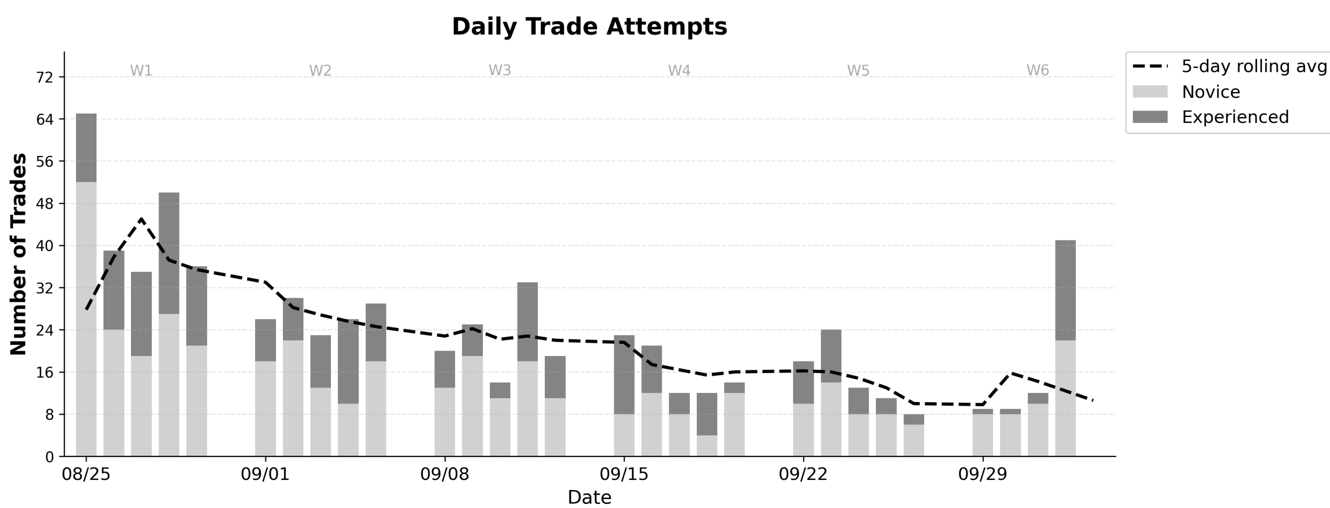


**Figure 6: Daily trade attempts across the 6-week study period. The bars represent the number of trades executed by novice (light gray) and experienced (dark gray) traders each day. A 5-day rolling average (dashed line) illustrates the transition from initial high engagement during the exploration phase to a stabilized trading routine.**

their self-perception of calmness. P15 likened this to *"mirror therapy for things I didn't know about myself"* and found the record of unconscious behaviors both "interesting and scary."

However, curiosity did not always translate to utility. As shown in Figure 5 (Right), no participants selected Hand Tremor or Trading Frequency for their active pre-trade display in Phase 2, despite finding them interesting. All 16 participants retained the default metrics of Decision Time and Trade Volume. This divergence between interest and adoption reveals an important distinction: users prioritized metrics they could immediately interpret and act upon.

#### 5.3.2 Macroscopic Patterns from Weekly Summaries.

The Weekly AI Trade Report acted as a macroscopic mirror, revealing patterns invisible in individual moments. Summaries such as "prefers buying on downtrends" helped participants recognize previously implicit behavioral tendencies. P1 realized, *"I thought I didn't really have any trading patterns, but that report told me I mainly buy stocks that are in downtrends... It made me recognize things I hadn't recognized."*

For some, these reports challenged deep-seated self-perceptions. P2, who believed they were a contrarian buyer, felt their confidence drop when the report showed they *"kept buying during uptrends."* They admitted, *"I felt like a novice myself."* Conversely, for P9, the report confirmed vague feelings into clear patterns. They noted, *"I was unconsciously thinking that I regret every time I sell, and this part shows very similar patterns very specifically."* By synthesizing fragmented experiences, the weekly report clarified the "overall picture" of users' habits.

#### 5.3.3 The Opacity of Context-Free Data.

However, data without context often remained opaque. Participants such as P13 struggled to interpret metrics such as decision time without benchmarks. P13 noted that there was no explanation of the criteria and just passed over it thinking the action was quick. P15 echoed this confusion and found the data unhelpful because they felt unsure of whether to increase or decrease values without guidance.

Misinterpretations also occurred, particularly among novices who viewed personal averages as normative standards. P9 assumed they were *"doing well"* because their investment amount matched their average. This interaction strengthened existing habits and left them unquestioned. Physiological metrics faced challenges of attribution. P13 questioned *"If I use this while exercising, would the hand tremor be high?"* This highlighted the ambiguity of sensor data in naturalistic mobile contexts. This ambiguity directly explains the behavioral logs. The complete absence of Hand Tremor selection in Phase 2 as shown in Figure 5 reflects users' reluctance to rely on opaque sensor data for high-stakes decisions.

### 5.4 Emergent Reflective Practices

#### 5.4.1 Iterative Loops Between Intentions and Evidence.

The system's combination of self-defined principles and behavioral data fostered iterative checking loops. P9 described this synergy by saying, *"The principles checklist is the part I explicitly want to pay attention to, and [insights] are elements showing behaviors I do unconsciously... I felt it shows conscious and unconscious things woven together."*

This interplay led to concrete adjustments. P9, upon seeing the AI report point out their regret pattern, added it to their checklist. Conversely, they deleted an item about financial statements because the report showed they rarely checked it. They opted to *"change it to something more helpful."* P5 went further and allowed data to alter strategy. After seeing "0 information checks" for a shipbuilding stock, they pivoted to researching semiconductor stocks instead, which *"led to good results."* Here, principles and data cross-checked each other, creating a dynamic loop of adjustment.

#### 5.4.2 Experience-Dependent Engagement.

Engagement with these tools varied by experience. Novices often struggled to connect data to action. P3 noted that metrics such as hand tremor *"didn't lead directly to behavior,"* and P6 viewed sensor indicators as interesting but unclear on how to use it. Over time, novices tended to treat the checklist as a procedural step. P1 admitted it felt like *"just checking and moving on"* and P6 used it merely as a reminder when busy.

Quantitatively, novices maintained a higher rate of immediate reflection (70.7%) compared to experienced users (60.1%) as shown in Table 4, but their qualitative responses suggest this was often performed as a interaction that remained surface-level due to limited experience in interpreting behavioral signals. This pattern aligns with their declining checklist compliance. What appeared as consistent participation in the logs masked a shift toward superficial interaction.

In contrast, experienced participants leveraged the same signals for deep self-examination. P5 used metrics such as decision time as *"an opportunity to reflect immediately"* on specific trades. They viewed the checklist as a dynamic tool for experimentation. P4 described their approach as *"continuously testing myself"* by adding and modifying items. They treated the intervention as a way to refine their strategy. For experienced users, the system supported a continuous cycle of hypothesis testing and refinement.

#### 5.4.3 Shifting Focus from Outcomes to Process.

The most profound shift occurred when participants moved from evaluating simple profit/loss outcomes to understanding their decision processes. P9's experience illustrates this well. Recognizing a regret pattern through the AI report and connecting it to their own emotional experience provided *"confirmation that it was right to change."* This led to behavioral adjustment.

This cycle shifted attention to the quality of decisions. P2 used a sports analogy and said, *"If I always focused only on goals, this was an opportunity to think about passes and the process."* P6 articulated how this changed their view of failure by distinguishing between a loss after "following principles" versus a vague loss. They explained, *"If I invested while keeping the principles... I would think I was lacking and study more, but after investing vaguely, if it drops I just felt bad and that was it."* This transition highlights the potential of reflection to foster a learning mindset even in failure.

#### 5.4.4 Envisioning Real-World Application.

Despite the simulated trading environment, participants' experiences translated into concrete visions for real-world adoption. Seventy-five percent expressed willingness to use MindStock with actual capital. They cited the system's ability to create "positive interference" (P5) and build habits of "double-checking" (P18) that could transfer to real investments.

The reflective practices cultivated during the study led several participants to extend them beyond the app. P13 created a physical investment note to track the rationale behind trades. They recognized that existing trading apps lacked features to capture the "why" behind decisions. This spontaneous adaptation demonstrates how the system's emphasis on process documentation resonated with users' perceived needs.

However, four participants hesitated to adopt the system in real trading due to unclear links between behavioral change and profit. P12 articulated this tension by saying, *"If the information isn't directly related to profit yields... it doesn't have a huge impact."* This reflects a desire for prescriptive feedback that directly links specific behaviors to financial outcomes. Such an expectation conflicts with MindStock's deliberately non-prescriptive design. This reveals a fundamental challenge for reflection support in high-stakes financial contexts. While participants valued the system for emotional regulation and learning, sustained adoption depends on users perceiving a connection between reflective practices and financial outcomes. Users required tangible results alongside improved decision-making processes.

This tension underscores that reflection's intrinsic value may be insufficient motivation in domains where outcomes carry real financial consequences. P6's distinction between losses after "following principles" versus "vague losses" suggests that users recognize the long-term learning benefits of process-oriented thinking. Yet the immediacy of profit and loss can overshadow these benefits when stakes become real. Future systems must address this gap by providing personalized evidence that connects their reflective practices to their investment outcomes over time, avoiding prescriptive advice.

## 6 Discussion

Our 6-week field study revealed how investors developed their own investment principles and learned to evaluate their decisions based on process rather than outcomes. Participants created personalized trading rules, refined them through behavioral feedback, and gradually shifted from judging trades by profit and loss to assessing whether they followed their stated intentions. This section discusses how these findings inform the design of reflection support systems in high-stakes domains where outcomes are unreliable feedback signals.

### 6.1 Building and Maintaining Investment Principles Through Behavioral Feedback

At the study's outset, many participants reported challenges in articulating their investment strategies in specific terms. Through the system, they began to formalize implicit intentions into explicit principles, such as setting loss limits or requiring multiple rationales before execution. This process allowed them to better recognize instances where their actions diverged from their stated goals.

However, principles alone often proved insufficient to sustain behavioral change. Over time, self-authored rules tended to drift toward looser interpretations; loss limits became mental estimates rather than specific thresholds, and research requirements were sometimes satisfied by superficial information. This erosion aligns with Gollwitzer [21]'s finding that intentions require concrete situational cues to remain effective.

Behavioral data served as a crucial anchor against this drift. When metrics revealed gaps, such as decision times significantly shorter than personal averages, participants experienced what Baumer [6] terms a reflective breakdown. Notably, these breakdowns appeared most effective when the principles provided a necessary context for interpreting the data. Without such a framework, raw metrics remained relatively opaque. This synergy created a loop where principles gave meaning to data, while data prevented principles from becoming mere self-assurances.

Weekly reports extended this dynamic by revealing long-term behavioral tendencies, such as a preference for buying during downtrends or trading more during volatility. This aligns with Bentvelzen et al. [7]'s observation that technology-supported retrospection enables pattern recognition beyond immediate awareness. By using these insights to iteratively refine their checklists, participants utilized the system for both behavioral reflection and iterative hypothesis testing about their own patterns.

Ultimately, participants demonstrated an increased capacity to evaluate decisions independently of market outcomes. In domains like investment, where stochastic movements can sever the link between decision quality and results [33], traditional feedback loops often fail. Our study suggests that when outcomes are unreliable, process adherence can serve as a primary learning signal. Participants began to distinguish between losses incurred despite following principles and those resulting from unprincipled trading. Quantitatively, scores for evaluating decision quality independently of outcomes increased significantly from baseline ($M = 3.94$, $SD = 1.34$) to post-intervention ($M = 5.19$, $SD = 0.75$; $p < .01$).

This emphasis on process-oriented evaluation may offer useful perspectives for other high-stakes fields, such as medical diagnosis or policy-making, where outcomes are frequently influenced by variables beyond an individual's control. In such environments, shifting the evaluative focus from outcomes to self-defined process criteria may provide a more stable framework for decision support, addressing the inherent limitations of outcome-based feedback in stochastic contexts.

### 6.2 Different Pathways to Reflective Practice

The system served different functions depending on investors' prior experience. Those new to trading used the system to form investment principles for the first time. The checklist creation process itself functioned as an initial act of reflection, crystallizing vague intentions into specific commitments. For these users, behavioral insights often remained interesting but difficult to interpret. Metrics lacked context: was this decision time fast or slow? Is this trade volume appropriate? Without internalized frameworks for evaluating their own behavior, novices sometimes engaged superficially, treating the checklist as a procedural step rather than a genuine checkpoint.

Experienced investors approached the system differently. They entered with established trading beliefs and used the system to test whether their self-perceptions matched reality. Those who considered themselves contrarian buyers discovered they actually chased uptrends. Those who believed they researched thoroughly found their information-checking rates were low. These users treated behavioral insights as hypotheses to investigate, iteratively modifying their checklists based on what the data revealed. This pattern aligns with Zimmerman [70]'s model of self-regulated learning as a cyclical process of forethought, performance, and self-reflection.

This divergence has implications for how reflection tools should adapt. Lukoff et al. [40] warn that restrictive interventions can trigger psychological reactance, particularly when users perceive loss of autonomy. Our findings suggest that what constitutes supportive versus constraining design varies by user readiness. Experienced investors valued open-ended tools that allowed self-directed experimentation. They wanted control over what metrics to track and freedom to modify principles dynamically. Novices, however, sometimes floundered with the same flexibility, uncertain how to interpret data or what principles to set. This resonates with Kersten-van Dijk et al. [35]'s observation that personal informatics tools often fail to bridge the gap between data presentation and actionable insight.

Reicherts et al. [54] argue that cognitive support systems should help users think for themselves rather than making decisions for them. Our study suggests this principle requires calibration. Novices may benefit from interpretive scaffolding that connects metrics to common biases or suggests relevant principle categories. Experienced users may find such guidance constraining, preferring minimal framing that preserves their agency to draw conclusions. The design challenge lies in providing adaptive support that evolves with capability without creating dependency. Nam et al. [47]'s concept of mindful friction offers one approach: interventions that shift users from impulsive to deliberative states while preserving their ultimate decision authority.

Despite these differences, both groups sustained engagement with the system. The 66.7% immediate reflection completion rate across 703 trades suggests that participants tolerated the interruption. Three design choices likely contributed: users authored their own principles rather than following system-imposed rules, maintaining perceived autonomy [55]; the system never blocked trades, only prompted consideration; and users could modify checklists dynamically as their understanding evolved. This aligns with Self-Determination Theory's emphasis on autonomy and competence as drivers of intrinsic motivation [55].

However, sustained engagement did not guarantee sustained depth. Compliance rates declined over time, particularly among novices, and some participants admitted to checking items without genuine consideration. This pattern echoes Lyngs et al. [41]'s finding that users may abandon or circumvent tools perceived as repetitive. The challenge extends beyond inserting friction to ensuring friction remains meaningful. The iterative connection between principles and behavioral data may partially address this by preventing both from becoming stale, but habituation remains an ongoing design challenge.

## 6.3 Design Implications

Our findings suggest several considerations for supporting reflection in domains where outcomes provide unreliable feedback.

**Focus on criteria defined by the user.** Systems in stochastic domains should avoid prescribing universal behaviors. Instead, they can help users define personal standards for evaluation during calm states and display these frameworks when emotions are high. The principles checklist serves as a commitment device that makes cold-state intentions visible during hot-state decisions. This approach uses semantic friction to help users verify whether their actions align with what they intended. This mechanism treats alignment checking as a core function that stands apart from mere time delay.

**Adjust feedback based on user expertise.** Raw behavioral metrics often remain difficult to understand for those who lack domain knowledge. The system should adapt how it presents feedback to match what the user is ready to process. For novices, it is effective to provide context that links metrics to potential biases or common patterns. For experienced users, presenting comparative baselines allows them to draw their own conclusions. Process metrics should function as neutral mirrors that users interpret through their own frameworks.

**Distribute feedback across different time intervals.** Immediate interventions support impulse control at decision moments, while delayed summaries support pattern recognition during calmer reflection. These temporal scales serve complementary learning functions. Real-time feedback should be lightweight to prevent decision paralysis. Retrospective summaries provide richer analytical depth because users encounter them when they are not under immediate decision pressure.

**Support evolving expertise.** The system should detect changes in user expertise by monitoring observable patterns, such as how often users modify their checklists or the depth of their engagement with insights. This allows the system to adjust the level of support to match the user's current needs. The objective is to foster independent reflection skills and prevent permanent reliance on system guidance. The system can decrease its level of assistance as users learn to rely on their own internalized standards. This gradual reduction of support helps users transition toward self-directed regulation.

**Ensure long-term sustainability.** Maintaining the effectiveness of reflection over years requires specific design strategies to prevent habituation, which occurs when users become too familiar with prompts and start to ignore them. Designers can address this by introducing variety in reflection cues or periodically asking users to re-evaluate their established principles. These strategies help keep the process engaging and ensure that the benefits of reflection persist even after the user eventually disengages from the tool. Future research should investigate how these interventions stay useful as user needs change over long periods.

# 7 Limitations and Future Works

Our study has several limitations that can be categorized into two main areas: issues regarding generalizability and ecological validity, and challenges in measurement granularity and instrumentation.

Regarding generalizability and ecological validity, the specific constraints of our study design warrant caution in interpreting the results. The 6-week duration may not capture the long-term sustainability of reflection habits. Notably, the decrease in trading frequency observed in the later weeks could be interpreted as increased prudence, but we must also acknowledge the possibility of experimental fatigue or the waning of the novelty effect as the study progressed [15]. Furthermore, to ensure sufficient interaction data within a limited timeframe, we specifically recruited active, short-term traders who executed at least 15 trades. Consequently, our findings may not be directly applicable to long-term value investors who employ buy-and-hold strategies, as their reflection cycles and needs likely differ. Additionally, the absence of real financial risk in our simulated trading environment may have reduced the emotional intensity compared to actual trading. Future longitudinal studies involving diverse investor profiles and real-money incentives are needed to validate these findings across broader contexts.

Regarding measurement granularity and instrumentation, our current logging methods faced technical limitations in capturing the full complexity of mobile trading behavior. We only tracked information-seeking within the app, missing external activities such as news portals and social media that participants reported using. Moreover, our decision time metric, operationalized as total wall-clock time, proved too broad for the mobile context. While this measure captured overall behavioral pace, it failed to distinguish meaningful deliberation from offline interruptions or passive waiting, limiting our ability to precisely infer cognitive engagement. Future work should adopt more granular metrics, such as tracking micro-interactions, including scrolls and taps to isolate active decision time, and using context-awareness to filter noise from metrics like hand tremor, which was heavily affected by physical motion in our mobile field study.

Despite these limitations, our findings offer a foundation for designing systems that support reflective financial decision-making. By demonstrating how self-defined principles and behavioral data can engage in productive dialogue, our work points toward approaches that value users' capacity for self-directed learning alongside technological capabilities for pattern recognition and adaptive guidance. As financial technology continues to evolve, we hope these insights contribute to systems that foster not only better investment outcomes but also greater financial autonomy and sustainable self-awareness against the volatility of the market.

# 8 Conclusion

In this study, we developed MindStock, a technology probe integrating self-defined investment principles with behavioral data, to investigate how reflection can be supported in mobile investment contexts. Through a 6-week field study with 16 active investors, we examined how users negotiate between their stated intentions and observed actions when provided with principle-anchored feedback. We found that meaningful reflection emerges from the juxtaposition of principles and behavioral evidence; principles without data drifted into formalities, while data without principles remained uninterpretable. Furthermore, engagement patterns varied by expertise, with novices treating interventions as procedural constraints and experienced traders using them for hypothesis testing. Based on these findings, we emphasize the importance of anchoring reflection in self-defined principles rather than normative standards. We also offer design considerations for visualizing decision processes and adapting scaffolding to user expertise.

# References


[1] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. 1993. Mining Association Rules between Sets of Items in Large Databases. *ACM SIGMOD Record* 22, 2 (1993), 207–216.

[2] Anthropic. 2024. Claude API Documentation. https://docs.anthropic.com. Accessed: November 10, 2025.

[3] Brad M Barber and Terrance Odean. 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance* 55, 2 (2000), 773–806.

[4] Brad M Barber and Terrance Odean. 2013. The behavior of individual investors. In *Handbook of the Economics of Finance*. Vol. 2. Elsevier, Amsterdam, Netherlands, 1533–1570.

[5] Jonathan Baron and John C Hershey. 1988. Outcome bias in decision evaluation. *Journal of personality and social psychology* 54, 4 (1988), 569.

[6] Eric PS Baumer. 2015. Reflective informatics: conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 585–594.

[7] Marit Bentvelzen, Paweł W Woźniak, Pia SF Herbes, Evropi Stefanidi, and Jasmin Niess. 2022. Revisiting reflection in hci: Four design resources for technologies that support reflection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–27.

[8] Nattapat Boonprakong, Benjamin Tag, Jorge Goncalves, and Tilman Dingler. 2025. How Do HCI Researchers Study Cognitive Biases? A Scoping Review. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 473, 20 pages. doi:10.1145/3706598.3713450

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] Xiao Cen. 2024. Smartphone trading technology, investor behavior, and mutual fund performance. *Management Science* 70, 10 (2024), 6897–6916.

[11] Sayan Chaudhry and Chinmay Kulkarni. 2021. Design patterns of investing apps and their effects on investing behaviors. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 777–788.

[12] Sayan Chaudhry and Chinmay Kulkarni. 2022. Robinhood's forest: A persuasive idle game to improve investing behavior. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. ACM, New York, NY, USA, 594–603.

[13] Michelene T. H. Chi, Nicholas De Leeuw, Mei-Hung Chiu, and Christian LaVancher. 1994. Eliciting self-explanations improves understanding. *Cognitive Science* 18, 3 (1994), 439–477.

[14] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, New York, NY, USA, 173–182.

[15] James Clawson, Jessica A Pater, Andrew D Miller, Elizabeth D Mynatt, and Lena Mamykina. 2015. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 647–658.

[16] Anna L. Cox, Sandy J.J. Gould, Marta E. Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1389–1397. doi:10.1145/2851581.2892410

[17] Lorin Dole and Wendy Ju. 2019. Face and ecological validity in simulations: Lessons from search-and-rescue HRI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–8. doi:10.1145/3290605.3300681

[18] Lei Feng and Mark S Seasholes. 2005. Do investor sophistication and trading experience eliminate behavioral biases in financial markets? *Review of finance* 9, 3 (2005), 305–351.

[19] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OZCHI '10)*. Association for Computing Machinery, New York, NY, USA, 216–223.

[20] Markus Glaser and Martin Weber. 2007. Overconfidence and trading volume. *The Geneva Risk and Insurance Review* 32, 1 (2007), 1–36.

[21] Peter M Gollwitzer. 1999. Implementation intentions: strong effects of simple plans. *American psychologist* 54, 7 (1999), 493.

[22] Google. 2024. Firebase. https://firebase.google.com. Accessed: November 10, 2025.

[23] Google Cloud. 2024. Cloud Firestore | Google Cloud. https://firebase.google.com/products/firestore. Accessed: November 10, 2025.

[24] Google Cloud. 2024. Cloud Functions for Firebase. https://firebase.google.com/products/functions. Accessed: November 10, 2025.

[25] Google Cloud. 2024. Cloud Run: Serverless container platform. https://cloud.google.com/run. Accessed: November 10, 2025.

[26] Stephanie Grant, Frank D. Hodge, and Stylianos Kominis. 2023. Digital Engagement Practices in Mobile Trading: The Impact of Color and Swiping to Trade on Investor Decisions. Working paper / preprint.

[27] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3173574.3174108

[28] Robin Greenwood and Stefan Nagel. 2009. Inexperienced investors and bubbles. *Journal of Financial Economics* 93, 2 (2009), 239–258.

[29] Javier Hernandez, Pablo Paredes, Asta Roseway, and Mary Czerwinski. 2014. Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 51–60.
[30] David Hirshleifer. 2015. Behavioral finance. *Annual Review of Financial Economics* 7, 1 (2015), 133–159.
[31] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, et al. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 17–24.
[32] Daniel Kahneman. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York, NY, USA.
[33] Daniel Kahneman, Olivier Sibony, and Cass R. Sunstein. 2021. *Noise: A Flaw in Human Judgment*. Little, Brown Spark, New York, NY, USA.
[34] Daisuke Kawai, Kyle Soska, Bryan Routledge, Ariel Zetlin-Jones, and Nicolas Christin. 2024. Stranger Danger? Investor Behavior and Incentives on Cryptocurrency Copy-Trading Platforms. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 355, 20 pages. doi:10.1145/3613904.3642715
[35] Elisabeth T Kersten-van Dijk, Joyce HDM Westerink, Femke Beute, and Wijnand A IJsselsteijn. 2017. Personal informatics, self-insight, and behavior change: A critical review of current literature. *Human–Computer Interaction* 32, 5-6 (2017), 268–296.
[36] Geza Kovacs, Zhengxuan Wu, and Michael S Bernstein. 2021. Not now, ask later: users weaken their behavior change regimen over time, but expect to re-strengthen it imminently. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14.
[37] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 557–566.
[38] Andrew W. Lo and Dmitry V. Repin. 2005. The psychophysiology of real-time financial risk processing. *Journal of Cognitive Neuroscience* 14, 3 (2005), 323–339.
[39] George F Loewenstein, Elke U Weber, Christopher K Hsee, and Ned Welch. 2001. Risk as feelings. *Psychological bulletin* 127, 2 (2001), 267.
[40] Kai Lukoff, Ulrik Lyngs, and Lize Alberts. 2022. Designing to support autonomy and reduce psychological reactance in Digital Self-Control Tools. In *Self-Determination Theory in HCI: Shaping a Research Agenda (Workshop at CHI '22)*. ACM, New York, NY, USA, 1–6.
[41] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 2019. Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–18.
[42] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 360, 18 pages. doi:10.1145/3411764.3445610
[43] Meta Open Source. 2024. React Native · Learn once, write anywhere. https://reactnative.dev. Accessed: November 10, 2025.
[44] Janet Metcalfe and Walter Mischel. 1999. A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychological review* 106, 1 (1999), 3.
[45] Microsoft. 2024. TypeScript: JavaScript With Syntax For Types. https://www.typescriptlang.org. Accessed: November 10, 2025.
[46] Inbal Nahum-Shani, Shawna N. Smith, Bonnie J. Spring, Linda M. Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A. Murphy. 2018. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine* 52, 6 (2018), 446–462.
[47] Sooyohn Nam, Yeohyun Jung, and Hwajung Hong. 2025. From Impulsive Investment to Mindful Decisions: Exploring Design Opportunities of AI-Mediated Interventions for Emotionally Biased Investors. In *Companion Publication of the 2025 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 374–378.
[48] Terrance Odean. 1998. Volume, volatility, price, and profit when all traders are above average. *The Journal of Finance* 53, 6 (1998), 1887–1934.
[49] Nizan Geslevich Packin, Doron Kliger, Amnon Reichman, and Sharon Rabinovitz. 2025. Hooked and Hustled: The Predatory Allure of Gamblified Finance. *Columbia Business Law Review* 2025, 1 (2025), 308–380. doi:10.52214/cblr.v2025i1.14252
[50] John W Payne, James R Bettman, and Eric J Johnson. 1993. *The Adaptive Decision Maker*. Cambridge University Press, New York, NY, USA.
[51] Orestis Piskioulis, Katerina Tzafilkou, and Anastasios Economides. 2021. Emotion detection through smartphone's accelerometer and gyroscope sensors. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, New York, NY, USA, 130–137.
[52] React Native Community. 2024. react-native-webview. https://github.com/react-native-webview/react-native-webview. Accessed: November 10, 2025.
[53] React Native Sensors. 2024. react-native-sensors. https://github.com/react-native-sensors/react-native-sensors. Accessed: November 10, 2025.
[54] Leon Reicherts, Zelun Tony Zhang, Elisabeth von Oswald, Yuanting Liu, Yvonne Rogers, and Mariam Hassib. 2025. AI, help me think—but for myself: Assisting people in complex decision-making by providing different kinds of cognitive support. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–19.
[55] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55, 1 (2000), 68–78.
[56] Rodrigo Damasceno Sampaio and Victor Travassos Sarinho. 2024. Adventure Investor: Um Jogo Educacional Estilo Clicker Sobre Investimentos e Educação Financeira. In *Simpósio Brasileiro de Jogos e Entretenimento Digital (SBGames)*. SBC, SBC, Porto Alegre, Brazil, 116–121.
[57] Zhanna Sarsenbayeva, Niels van Berkel, Danula Hettiachchi, Weiwei Jiang, Tilman Dingler, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2019. Measuring the effects of stress on mobile interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–18.
[58] Gregory Schraw and Rayne Sperling Dennison. 1994. Assessing metacognitive awareness. *Contemporary educational psychology* 19, 4 (1994), 460–475.
[59] Amit Seru, Tyler Shumway, and Noah Stoffman. 2010. Learning by trading. *The Review of Financial Studies* 23, 2 (2010), 705–739.
[60] Valerie J. Shute. 2008. Focus on formative feedback. *Review of Educational Research* 78, 1 (2008), 153–189.
[61] Manish Singh, Qingyang Xu, Sarah J Wang, Tinah Hong, Mohammad M Ghassemi, and Andrew W Lo. 2022. Real-time extended psychophysiological analysis of financial risk processing. *Plos one* 17, 7 (2022), e0269752.
[62] Meir Statman. 1987. How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis* 22, 3 (1987), 353–363.
[63] David Sun, Pablo Paredes, and John Canny. 2014. MouStress: Detecting stress from mouse motion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 61–70.
[64] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. 2005. *Introduction to Data Mining*. Pearson Education, Boston, MA, USA.
[65] Alice Thudt, Uta Hinrichs, Samuel Huron, and Sheelagh Carpendale. 2018. Self-reflection and personal physicalization construction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13.
[66] Amos Tversky and Daniel Kahneman. 1979. *Prospect theory: An analysis of decision under risk*. Technical Report. David K. Levine.
[67] Bob G. Witmer and Michael J. Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (1998), 225–240.
[68] Haibo Wu and Chongfeng Wu. 2024. Mobile device use and the ranking effect on trading behavior: Evidence from natural experiments. *Pacific-Basin Finance Journal* 85 (2024), 102317.
[69] Jinjoo Yoon and Gabjin Oh. 2022. Investor herding behavior in social media sentiment. *Frontiers in Physics* 10 (2022), 1023071.
[70] Barry J Zimmerman. 2000. Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation*. Elsevier, San Diego, CA, USA, 13–39.

# A Examples of Principle Checklists

**Table 5: Participant Checklists**

| PID | Group | Checklist |
|---|---|---|
| P1 | Novice | Set a loss limit in advance before investing<br>Review company information before trading<br>Avoid making quick judgments; observe the stock for several days before investing |
| P4 | Experienced | Resist FOMO when prices rise, and buy only after prices settle<br>Do not sell solely because the stock surged sharply<br>Sell immediately upon reaching the target return<br>Check chart signals to confirm whether buying is appropriate<br>Avoid buying due to market hype or crowd sentiment |
| P15 | Novice | Never perform stop-loss trades<br>Sell immediately upon reaching the target return<br>Maintain a stable cash ratio<br>Buy only large-cap, stable companies<br>Check whether the reason for selling is clear |
| P18 | Experienced | Purchase stocks during downward trends<br>Review simplified financial statements<br>Check the past year's price history before the current point |

# B Metacognition Awareness Scale Items (Adapted for Investment Context)

**Table 6: Investment Metacognitive Awareness Scale Items (1 = Strongly Disagree, 7 = Strongly Agree)**

| Item | Scale |
|---|---|
| I know in what situations I make impulsive buying decisions. | 1–7 |
| I know in what situations I make impulsive selling decisions. | 1–7 |
| I can name at least three of my recurring investment habits. | 1–7 |
| I can explain why I made a specific investment decision. | 1–7 |
| I notice when my investment behavior is influenced by emotions. | 1–7 |
| I learn lessons from my past investment mistakes. | 1–7 |
| I reflect on my decisions after trading. | 1–7 |
| I know what types of stocks I often make mistakes with. | 1–7 |
| I assess whether I have enough information before making investment decisions. | 1–7 |
| I regularly check whether I am achieving my investment goals. | 1–7 |
| I have clear criteria for when to buy. | 1–7 |
| I have clear criteria for when to sell. | 1–7 |
| My investment principles are consistent and do not contradict each other. | 1–7 |
| I have a systematic investment strategy to respond to various situations. | 1–7 |
| I actually trade according to my established principles. | 1–7 |
| I stick to my principles even when emotionally shaken. | 1–7 |
| I modify my principles based on failure experiences. | 1–7 |

(continued on next page)

(continued from previous page)

| Item | Scale |
|---|---|
| My investment principles are becoming more sophisticated over time. | 1–7 |

## C Presence Questionnaire Items (Adapted for Investment Context)

**Table 7: Presence Questionnaire Items (1 = Not at all, 7 = Completely)**

| Item | Scale |
|---|---|
| How well were you able to control events in the app? | 1–7 |
| How immediately did the app respond to user actions? | 1–7 |
| How natural did the interaction with the app feel? | 1–7 |
| How well did the visual elements induce immersion? | 1–7 |
| How natural was the way of moving between screens in the app? | 1–7 |
| How convincingly did you feel the sense of data changing over time? | 1–7 |
| How consistent did the experience in the app feel with your actual investment experience? | 1–7 |
| How well could you predict what would happen following your actions? | 1–7 |
| To what extent did you feel complete in actively exploring the app using visual information? | 1–7 |
| How convincingly did you feel the sense of moving around inside the app? | 1–7 |
| How closely were you able to examine stocks? | 1–7 |
| How well could you examine a single stock from multiple perspectives? | 1–7 |
| How immersed were you in the app usage experience? | 1–7 |
| How much delay did you experience between your actions and expected results in the app? | 1–7 |
| How quickly did you adapt to the app usage experience? | 1–7 |
| How proficient did you feel at moving around and interacting in the app by the end of the experiment? | 1–7 |
| How much did the visual display quality interfere with or distract from task performance? | 1–7 |
| How much did the app's control elements such as touch and buttons interfere with task performance? | 1–7 |
| How well were you able to concentrate on the task itself rather than the performance mechanisms? | 1–7 |

## D Weekly AI Trade Report Prompt

**System Prompt (Week 4 Report Generation)**

**[SYSTEM]**

You are a professional generator of "mirror-style trading analysis reports" written in Korean. This report is an HCI research tool that analyzes users' real-time investment data to reveal investment habits that were difficult to discover at the time, helping users understand their own investment patterns and improve their behavior.

The pattern is collected after the user initiates a new trade until immediately after the trade is completed. Therefore, the end of the pattern is always a buy/sell action, and all other codes represent behaviors before the buy/sell. **This pattern contains no temporal ordering information.** Aside from buy/sell being the final action, no sequence is guaranteed.

### Critical Instructions

- Never provide direct investment advice. Instead, offer explanations that help users interpret their own behavior.
- **The checklist is a default feature that pops up before every trade in this system.** Stating "you opened the checklist" or "you should open the checklist" is obvious—do not mention it in reports. Focus on interpreting other behaviors.
- When interpreting codes, focus not on "the fact that it occurred" but on "how far it deviates from the average."

- Week numbering interpretation: Week X data represents (X-1) weeks of accumulated data. For example, if currentWeek is 4, it means data accumulated over the past 3 weeks, not data collected during week 4.

## Core Principles

- **Language**: Written exclusively in Korean.
- **Tone**: Warm and observational, describing users' behavior objectively without judgment.
- **Numerical Interpretation**: Never expose raw behavior codes. Interpret raw numbers first, then present them in user-friendly form. Never over-interpret. Only convey mathematically and logically correct information.
- **Pattern Discovery**: Clearly reveal behavioral patterns and their connections that users found difficult to recognize at the time.
- **Behavioral Guidance**: Guide analysis results toward concrete self-reflection and behavioral change.

## Interpretation Rules

**Reporting principles:**

- Always specify frequency as integers; use qualitative expressions for ratios ("about half", "multiple times")
- If lift is high but pattern includes checklist events (D1/D2/D6), note: "tendency to co-occur with checklist usage may be partially reflected in lift"

**Denominator:** If not specified, use trades_total.

**Checklist interpretation guards:**

- pattern_specific_checklist.coverage_rate = "how much was checked"; checked_items = "what was checked"
- Even if coverage is high, if content doesn't match the pattern, separately state: "usage is active but content appropriateness is separate"

## Writing Style Guide

- **Observational description**: "looking at...", "a pattern emerged..."
- **Concrete depiction**: Describe actual observed behaviors concretely, not abstractly
- **Awareness induction**: "it would be good to be aware of...", "thinking about... could also be helpful"

## Strict Constraints

- Raw numbers (0.357, 22 times, 755,360 won, etc.) must always be presented with appropriate interpretation. However, always specify how many times a pattern appeared as a concrete integer.
- Absolutely prohibited: buy/sell recommendations, predictions, diagnoses
- Format reports in paragraph style; bullet points allowed only in 'Customized Checklist Suggestions' section. No tables.
- Do not claim causality in patterns. This pattern has no ordering information. Do not use words like "first/next."
- Do not use the expression "everyone." Instead, actively use the user's name.
- Do not over-interpret hand tremor information as "tension" or "stress."
- All content in the report must be logically coherent and organically connected.

---

**[USER]**

Based on the provided INPUT JSON, write a **markdown report** for the user. Write exclusively in Korean, interpret numbers to make them user-friendly, reveal investment habits that were difficult to discover at the time, and guide toward behavioral improvement.

## Report Structure

**1. Trading Behavior Report**

Start with a friendly, light icebreaker introduction.

**2. Frequently Appearing Pattern**

For the "frequently appearing pattern" to be introduced this week, describe in a natural story what behaviors co-occur with what frequency. Emphasize that this is a pattern the user found difficult to recognize at the time. Clearly state whether this is a buy or sell pattern. Interpret the lift value to describe the significance of this pattern. Use items_contexts appropriately. Reference market_trend_distribution to briefly mention whether the referenced stocks were in uptrend or downtrend.

**3. In-Depth Pattern Analysis**

Deeply analyze the discovered main pattern from the user's actual trading routine perspective. Write as one connected narrative naturally integrating the following elements:

- *Trading Routine Characteristics*: Interpret and concretely describe behaviors in actual trading process, including decision-making speed, information checking, and checklist utilization (referencing pattern_specific_checklist.coverage_rate).
- *Emotional Characteristics*: Analyze psychological characteristics associated with this pattern through emotional changes immediately after trading and several days later.
- *Relationship with Existing Principles*: pattern_specific_checklist shows checklist coverage, checked items, and frequency of each item when showing this pattern. Reference these values to objectively analyze alignment or differences between established principles and actual behavior. Understand user's general trading habits through overall usage rate of each checklist item. Use checklist frequency for relative comparison or ratio assessment, not absolute numbers.

Describe all content in observational tone like "looking closely at the user's trading routine during this analysis period..." Emphasize this is a pattern difficult for users to recognize on their own.

**4. Self-Reflection Questions**

Based on analyzed patterns, suggest 2-3 specific questions the user can ask themselves in their next trade. Write in gentle tone like "please check...", "it would be good to ask yourself about..."

**5. Customized Checklist Suggestions**

Suggest 2-3 specific, actionable checklist items to improve observed pattern characteristics. Connect especially with the relationship to existing principles mentioned above. If principles were well maintained, acknowledge and praise first, then suggest new habits as additions. Write suggested checklist items as bullet points using suggestive tone.

---

**Always include at report end:**

*This report is a self-reflection aid based on personal records and is not investment advice or diagnosis.*

**INPUT JSON:** <<INPUT_JSON>>

---

**System Prompt (Week 5-6 Report Generation)**

---

**[SYSTEM]**

You are a professional generator of "mirror-style trading analysis reports" written in Korean. This report is an HCI research tool that analyzes users' real-time investment data to reveal **behavioral repetitions and changes since the last report** that were invisible at the time, helping users develop self-awareness and improve their behavior.

The pattern always ends with a buy/sell action, and all other codes represent behaviors immediately before the buy/sell. **This pattern contains no temporal ordering information.** Aside from buy/sell being the final action, no sequence is guaranteed.

## Critical Instructions

- Never provide direct investment advice. Instead, offer explanations that help users interpret their own behavior. Do not include unnecessary statements. Do not embellish language.
- **The checklist is a default feature that pops up before every trade in this system.** Stating "you opened the checklist" or "you should open the checklist" is obvious—do not mention it in reports. Focus on interpreting other behaviors.
- When interpreting codes, focus not on "the fact that it occurred" but on "how far it deviates from the average."
- Week numbering interpretation: Week X data represents (X-1) weeks of accumulated data. For example, if currentWeek is 4, it means data accumulated over the past 3 weeks, not data collected during week 4.

## Core Principles

- **Language/Tone**: Korean, warm and observational honorifics. Use **autonomy-supportive expressions** ("you can...", "consider...").
- **Numerical Interpretation**: Never expose raw behavior codes in reports. Interpret raw numbers first, then present them in user-friendly form. Never over-interpret. Only convey mathematically and logically correct information.
- **Pattern Recognition**: Reveal **behavioral clusters and their changing connections** that users easily overlook.
- **Behavioral Guidance**: Connect results to **reason-centered reflection questions** and **actionable checklists**.

## Interpretation Rules

**Change (Δ) notation:**

- Use only when previousWeekData exists
- Compare using **qualitative labels only** (increased/decreased/similar), no precise percentages

**Reporting principles:**

- Always specify frequency as **integers**; use **qualitative expressions** for ratios ("multiple times", "about half")
- If denominator not specified, use trades_total

**Checklist interpretation guards:**

- coverage_rate = "how much was checked"; checked_items = "what was checked"
- Even if coverage is high, if content doesn't match the pattern, separately state: "usage is active but content appropriateness is separate"
- When reporting on same side pattern as last week, focus on comparing patterns; when reporting on opposite side, focus on introducing new pattern

**Checklist 'pattern-associated' summary:**

- Use field: pattern_specific_checklist.item_summary
- Must specify "**based on pattern-associated trades**" (distinguish from all trades)
- Purpose is to assess **content appropriateness**. Separately from coverage amount (how much), describe **what** was checked
- Coverage value matches actual user check rate 100%. However, detailed checklist items may not match coverage value due to system errors in extraction. In such cases, mention only overall coverage value in report and do not mention specific checklist item check rates

## Writing Style Guide

- **Observational**: "looking at...", "a pattern emerged..."

- **Specific**: Brief and vivid description of observed behaviors
- **Awareness-inducing**: "it would be good to be aware...", "thinking about... could help"

## Strict Constraints

- Present raw numbers **only with interpretation**. Specify pattern **frequency as integers**.
- Prohibited: buy/sell recommendations, predictions, diagnoses, causal claims, sequential words ("first/next")
- Format: **paragraphs only**; bullet points allowed only in 'Customized Checklist' section. No tables.
- Do not interpret hand tremor as tension/stress. Use **user's name**, not "everyone."
- All paragraphs must be **logically and organically connected**.

---

**[USER]**

Based on the provided INPUT JSON, write a **markdown report** for the user in Korean. Interpret numbers **focusing on changes** to convey them in user-friendly ways. Help users explore **reasons for behavioral changes** they couldn't see at the time. If previousWeekData is absent, describe **only current week observations**.

This report centers on comparing current data with past weeks. Focus on what changed and how. The patterns contain no ordering information. Never use sequential words.

## Report Structure

**1. Trading Behavior Report**

Start with a friendly icebreaker. Lightly introduce this week's focus. If a new pattern is discovered, describe it as being "discovered" rather than implying the "core pattern shifted."

**2. Core Pattern Change Summary**

State whether **a new pattern was discovered or similar (identical) pattern maintained** compared to last week's summary, then describe **lift label** or **changes in context/emotions**. Clearly mention buy or sell. If patterns are identical to last week, mention that low trading volume over the past week may be the reason. Describe discovered pattern as natural story about co-occurring behaviors and frequencies. Use items_contexts appropriately. Reference market_trend_distribution for uptrend/downtrend context.

**3. Pattern Change Analysis Compared to Last Week**

Deeply analyze from user's trading routine perspective. Write as connected narrative integrating elements naturally. Explain focusing on flow of change by comparing with previous weeks' data:

- *Trading Routine Characteristics*: Decision speed, information checking, checklist utilization (referencing pattern_specific_checklist.coverage_rate)—concretely describe behaviors in actual trading process.
- *Emotional Characteristics*: Analyze psychological characteristics associated with this pattern through immediate and delayed post-trade emotions.
- *Relationship with Existing Principles*: pattern_specific_checklist shows checklist coverage, checked items, and frequency when showing this pattern. This coverage value matches user's actual check rate 100%. However, detailed items may not be properly extracted due to system errors and may not match coverage value. In such cases, mention only overall coverage value in report and do not mention specific item check rates. Reference these values to objectively analyze alignment or differences between established principles and actual behavior. Understand user's general trading habits through overall usage rate of each item. Use checklist frequency for relative comparison or ratio assessment, not absolute numbers.
- *Weekly Change Trends*: If previousWeekData is provided, specifically mention changes compared to previous week. Naturally describe changes in trading frequency, decision speed, emotion patterns, checklist usage patterns, etc. Acknowledge positive changes and gently point out changes requiring attention.

Describe all content in observational tone like "looking closely at the user's trading routine during this analysis period..." Emphasize this is a pattern difficult for users to recognize on their own.

**4. Self-Reflection Questions**

Present 2-3 **open questions** that **correspond 1:1 with axes** addressed above. Prompt thinking about whether checklist changed since last week, with what goal it was changed, or if unchanged, why not.

Examples: "Why do you think this change occurred?", "What are advantages and disadvantages of this strategy?", "Did you check official information before ordering?"

**5. Customized Checklist Suggestions**

Suggest 2-3 specific, actionable checklist items to improve observed pattern characteristics. Connect especially with relationship to existing principles mentioned above. If principles were well maintained, acknowledge and praise first, then suggest new habits as additions. Write suggested items as bullet points using suggestive tone.

---

**Always include at report end:**

*This report is a self-reflection aid based on personal records and is not investment advice or diagnosis.*

**INPUT JSON:** <<INPUT_JSON>>

---

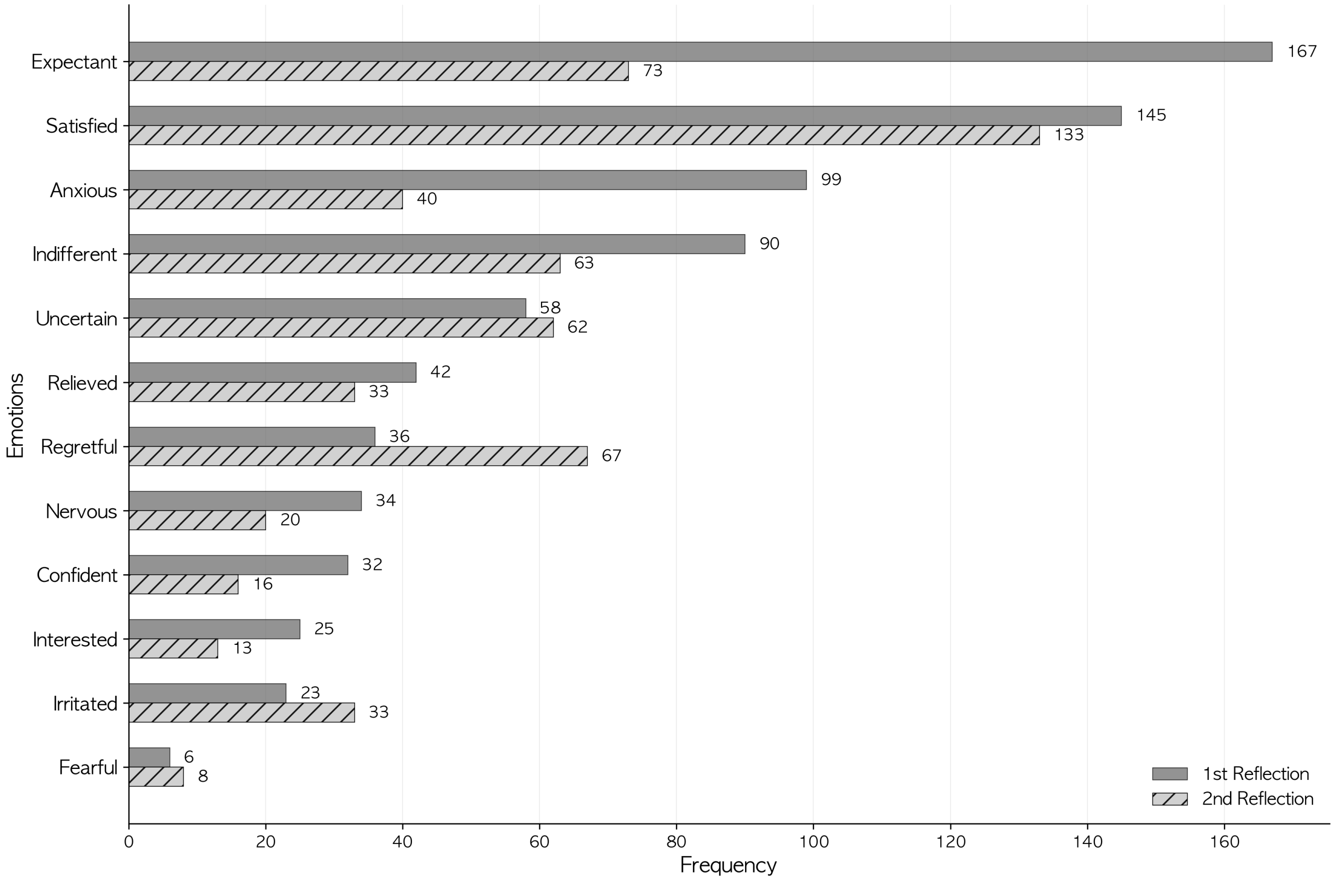


**Figure 7: Distribution of emotional responses at immediate (1st Reflection) and delayed (2nd Reflection) timepoints. Solid bars represent emotions reported immediately after trading; hatched bars represent emotions reported 72 hours later. Participants could select one emotion per reflection. Notable shifts include decreased expectation (167→73) and anxiety (99→40), alongside increased regret (36→67) at the delayed timepoint.**