Supplementary Material E. Canonical vs. PCA Items and Scale Architecture

For the purposes of the PCA validation study, the canonical item list was built as a theoretical "content universe" for each latent construct. We used a mainly deductive strategy from well-known scales where they exist, and an inductive, corpus-based strategy only where no such scales are available. This follows current guidance that scale development should start with precise concept definitions and systematic domain mapping before any statistical refinement (Podsakoff et al., 2016; Hinkin, 1995; DeVellis, 2016; Boateng et al., 2018; Morgado et al., 2017). The canonical item pool is therefore a reference point against which the brevity and coverage of our Supplementary material B instrument can be compared in terms of definitional correspondence and distinctiveness (Colquitt et al., 2019).

For the Theory of Planned Behavior, we treated Ajzen's original operationalisation and later measurement guidelines as the canonical source. Ajzen specifies that attitudes toward a behavior should be measured with bipolar evaluative adjectives such as "good–bad," "harmful–beneficial," "useful–useless," and similar semantic differentials, always tied to a specific target behavior (e.g., "For me, using an LLM when planning trades is bad–good"). Subjective norms are represented through injunctive and descriptive items of the form "Most people who are important to me think that I should [perform the behavior]" and "Most people like me [perform the behavior]," while perceived behavioral control/self-efficacy is captured by items such as "I am confident that I can [perform the behavior]" and "For me, [performing the behavior] would be very difficult–very easy" (Ajzen, 1991, 2006). In our canonical list we therefore combined the full set of attitude, norm, and control templates that Ajzen explicitly recommends as standard operationalisations for TPB constructs, treating them as the complete reference universe for ATT, SN, and self-efficacy content.

For the Technology Acceptance Model, we defined the canonical pool as the original six perceived usefulness (PU) and six perceived ease of use (PEOU) items introduced by Davis (1989). These include statements such as "Using the system improves my job performance," "Using the system increases my productivity," and "Learning to operate the system is easy for me," which have been repeatedly replicated and slightly adapted across TAM2/TAM3 and related models. Because Davis provides a complete, psychometrically validated inventory of PU and PEOU content, we did not reconstruct TAM items from definitions alone; instead, our canonical list is simply the union of the original PU and PEOU items, with minimal rewording of "system" into "LLM-based tools for trading" to preserve construct meaning while adapting the behavioural referent to our context.

For financial risk tolerance, the situation is similar: Grable and Lytton's (1999) 13-item instrument is one of the few multi-scenario scales explicitly developed and validated as a comprehensive risk-tolerance measure, and no widely accepted short form with full psychometric equivalence has emerged. The canonical list therefore consists of all 13 original items, which combine self-descriptions ("How would your best friend describe you as a risk taker?") with behavioural vignettes involving lotteries, investment choices, windfalls, and loss framing. These items jointly cover general risk comfort, investment versus speculative risk, and willingness to accept short-term losses for long-term gain. We use this full inventory as the reference universe for evaluating the facets captured by our shorter, scenario-based risk-tolerance block.

For trust in automation, both item-level and short-form guidance are available. We defined the canonical item universe as Jian, Bisantz and Drury's (2000) 12-item scale, which differentiates distrust (e.g., "The system is deceptive," "I am suspicious of the system's actions") from trust (e.g., "I am confident in the system," "The system is reliable," "I can trust the system"). This long form has been repeatedly used as

the foundational operationalisation of trust in automated systems and, more recently, in AI contexts, where it has also been subjected to comprehensive psychometric evaluation and reduction to short forms such as the three-item S-TIAS (e.g., McGrath et al., 2025). Our canonical list therefore includes all 12 Jian et al. items and notes the existence of validated three-item cores as a guide for constructing brief trust blocks in specific domains.

Perceived Cognitive Assistance (PCA) is different from these constructs because no prior scale exists. For PCA, the canonical item list had to be built inductively from the conceptual definition and an extensive qualitative corpus rather than imported from an existing questionnaire. Following concept-definition guidelines that stress explicit articulation of core attributes and careful separation from neighbouring constructs (Podsakoff et al., 2016), we first defined PCA as the trader's perception that an LLM allows them to organise, process, and translate complex trading information into executable strategies more effectively, without collapsing into general usefulness, ease of use, trust, or generic self-efficacy. We then used reflexive thematic analysis of three data sources – semi-structured interviews with retail traders, curated YouTube conversations, and robo-advisor/fintech studies – to identify recurring ways in which traders experience perceived cognitive help, including structural scaffolding, cognitive load relief, error checking, volatility navigation, and autonomy boundaries. These themes were translated into an intentionally overcomplete item pool (e.g., "Using an LLM helps me break complex trades into manageable steps," "With LLM help I can handle more information without feeling overloaded," "LLM support enables me to translate my market view into a precise, executable trade plan"), which constitutes the canonical PCA item universe for this study.

In sum, the canonical list collates, for each construct, either (a) the full item set from well-known, repeatedly validated scales where such instruments exist (TPB, TAM, Grable–Lytton risk tolerance, Jian et al. trust in automation) or (b) a theoretically grounded, qualitatively derived item universe where the construct is novel (PCA). This mixed approach — deductive for established constructs, inductive for PCA — means our shorter Supplementary material B items can be checked against a clear target set of content. This follows current recommendations for concept definition, item generation, and content validation in the organisational and behavioural sciences (Podsakoff et al., 2016; Boateng et al., 2018).

Across the four established constructs (TPB, TAM, financial risk tolerance, trust in automation) we started from a canonical pool of approximately 57 items drawn from standard instruments in social psychology, IS acceptance, household finance, and human–automation interaction (Ajzen, 1991; Davis, 1989; Grable & Lytton, 1999; Jian et al., 2000; McGrath et al., 2025). From this universe we retained 28 legacy items (13 TPB, 6 TAM, 6 risk tolerance, 3 trust), corresponding to roughly 49% of the canonical content, and added a 16-item pool for the novel PCA construct, yielding 44 items in total. This selection balances content coverage and survey length while keeping each construct close to its theoretical definition.

For TPB, we preserve a full 13-item block: a five-item, behaviour-specific attitude index using Ajzen-style bipolar evaluative anchors; three subjective-norm items covering valued others, expert peers and close social circle; and five trading self-efficacy items rated on a 0–100 confidence scale rather than agreement Likert. This is consistent with Ajzen's recommendation that TPB measures be tailored to the focal behaviour and with guidance that self-efficacy be assessed via graded confidence judgments in domain-specific capabilities, especially for complex multi-step behaviours such as trading (Bandura, 1997, 2006).

For TAM, we reduce Davis's original 6+6 PU/PEOU battery to three usefulness and three ease-of-use items that are tightly focused on the trading context: decision quality, planning quality and overall utility

for PU; learnability, clarity, and controllability for PEOU. Prior TAM extensions and short-form usability scales indicate that such brief subscales can retain acceptable reliability when they preserve the core conceptual facets, as in the UMUX/UMUX-LITE work where two-item usefulness/usability composites achieved internal consistencies above 0.80 (Lewis et al., 2013). If pilot reliability for PU or PEOU falls below 0.75, one additional item per construct ("enables efficient analysis"; "overall easy to use") can be added without disrupting the conceptual mapping.

For financial risk tolerance, we replace the earlier abstract Likert items with a six-item scenario-based short form directly sampled from the Grable and Lytton (1999) instrument. The selected items (friend-description, windfall allocation, meaning of "risk," best/worst return profile, inheritance allocation, and stock-comfort) preserve the original multidimensional structure, covering general risk comfort, investment versus speculative risk, and cognitive framing of risk as loss, uncertainty, or opportunity. This keeps comparability with the widely used G&L scoring logic while reducing respondent burden relative to the full 13-item scale and avoiding construct under-representation.

Trust in automation is measured with a three-item block that mirrors the high-loading items from the Trust in Automation Scale (TIAS) and its recently validated three-item short form (S-TIAS), both of which show very high internal consistency when used to assess trust in AI systems. Our items explicitly target confidence in the LLM as a starting point, perceived predictability of its behaviour, and comfort in acting on LLM-assisted plans after personal review. These "appropriate-use" qualifiers (e.g., "as a helpful starting point," "after I review it") are included to prevent blind-reliance interpretation, in line with trust-calibration guidance (Lee & See, 2004).

Finally, PCA (Perceived Cognitive Assistance) is operationalised via an intentionally over-specified 16-item pool that covers six qualitatively derived themes (structural support, cognitive load handling, error checking, volatility navigation, transparency, and autonomy/breadth of exploration). The pool was built from the triangulated corpus described above (interviews, YouTube retail-investor content, robo-advice/fintech studies) following current recommendations for concept and content specification in scale development. Study 2 establishes definitional correspondence and distinctiveness via naïve-judge content validation (Supplementary material A; results in Supplementary material D), prior to any factor analysis. In subsequent trader-survey waves, the PCA block is evaluated alongside adjacent constructs (PU, PEOU, trust, and trading self-efficacy) using exploratory and confirmatory factor-analytic models, with convergent and discriminant validity assessed using standard criteria (e.g., AVE, HTMT, and nested model comparisons; Henseler et al., 2015; Fornell & Larcker, 1981).

Note on scope: Study 2 (naïve-judge content validation) used only the 16 PCA candidate items plus four comparator filler items, as documented in Supplementary material A. Table S18 documents the canonical-versus-retained logic for the full multi-construct instrument used in later quantitative waves.

Table S18. Scale Comparison: Canonical vs. PCA Items.

Theory / Theme	Seminal Works	Canonical Scale Items (ALL)	Retained Survey Items (Supplementary material B)	Rationale for Retained Block
1. TPB (Theory of Planned Behavior)	Ajzen (1991, 2006)	ATTITUDE (5–7 items): harmful/beneficial; worthless/valuable; useless/useful; foolish/wise; bad/good; unpleasant/pleasant;	ATTITUDE (5 items – bipolar evaluative, behavior-specific): ATT1: harmful … beneficial; ATT2: unwise … wise; ATT3: bad … good; ATT4: worthless … valuable; ATT5: anxiety-provoking … confidence-building	ATTITUDE: 5-item index with bipolar anchors is fully consistent with TPB measurement practice.

		boring/interesting; unenjoyable/enjoyable; dull/exciting; stressful/relaxing SUBJECTIVE NORMS (3–5 items): People important to me think I should…; It is expected of me that…; People I value would approve…; Most people like me do…; Most [reference group] do… PBC/SELF-EFFICACY (5–6 items): I am confident I can…; For me to [behavior] would be (impossible/possible); For me to [behavior] would be (difficult/easy); How much control do I have…; Whether I [behavior] is up to me; I have complete control over…	SUBJECTIVE NORMS (3 items): SN1: People whose opinions I value think I should…; SN2: Most traders I respect approve…; SN3: People close to me support… SELF-EFFICACY (5 items, 0–100 scale): TSE1: organise complex; TSE2: find alternatives; TSE3: keep to plan under pressure; TSE4: judge when need help; TSE5: evaluate and learn	SUBJECTIVE NORMS: 3 items cover general important others, domain experts, close circle. SELF-EFFICACY: Domain-specific 0–100 scales tend to produce higher reliability and better predictive validity than generic agreement formats, particularly when the behavior is complex and multi-step (Bandura, 1997, 2006).
2. TAM (Technology Acceptance Model)	Davis (1989)	PERCEIVED USEFULNESS (6 items): Accomplish tasks more quickly; Improve job performance; Increase productivity; Enhance effectiveness; Make it easier to do job; I would find [system] useful PERCEIVED EASE OF USE (6 items): Learning to operate would be easy; Easy to get [system] to do what I want; Interaction clear and understandable; Flexible to interact with; Easy to become skillful; I would find [system] easy to use	PERCEIVED USEFULNESS (3 items): PU1: Improves my trading decisions; PU2: Increases quality of trade planning; PU3: Overall useful for trading tasks PERCEIVED EASE OF USE (3 items): PEOU1: Learning to use LLM has been easy; PEOU2: Interacting is clear/understandable; PEOU3: Easy to get LLM to do what I need	Both PU & PEOU: 3 items each. Strong precedent: TAM2/TAM3 (Venkatesh & Davis, 2000) used 4 items; UMUX-LITE validated 2 items (α = .82–.86; Lewis et al., 2013). Our 3 items capture: (PU) decision quality, process quality, global utility; (PEOU) learnability, clarity, controllability. If pilot $\alpha < .75$, add PU: “enables efficient analysis” and PEOU: “overall easy to use.”

3. Financial Risk Tolerance (scenario-based)	Grable & Lytton (1999)	ALL 13 ITEMS (full text): How would friend describe you as risk taker? (4 options) | TV game show choice? | Vacation after job loss? | $20K windfall invest where? | Comfort investing in stocks? | "Risk" = ? | Bonds→hard assets scenario? | Best/worst returns preference? | Given $1K, gain choice? | Given $2K, loss choice? | $100K inheritance invest? | $20K allocation? | Gold mine venture?	6 items (scenario-based; categorical response options as in Grable & Lytton, 1999): RT1 (Item 1): "How would your best friend describe you as a risk taker?" [4 options] RT2 (Item 4): "If you received a $20,000 windfall, how would you invest it?" [bank/bonds/stocks] RT3 (Item 6): "To me, the word 'risk' means…" [loss/uncertainty/opportunity/thrill] RT4 (Item 8): "Which pattern of best and worst returns would you prefer?" [4 graded options] RT5 (Item 11): "If you inherited $100,000 today, how would you invest it?" [savings/balanced/stocks/commodities] RT6 (Item 5): "How comfortable are you with the idea of investing in stocks?" [not at all/somewhat/very]	A scenario-based short form directly sampled from the validated Grable & Lytton (1999) instrument. The 6 core items (1, 4, 6, 8, 11, 5) preserve the original multidimensional structure (general risk comfort, investment vs. speculative risk, gain–loss framing, meaning of "risk") while keeping survey length manageable. Scoring follows G&L's original risk-point scheme, keeping comparability with prior work.
4. Trust in Automation	Jian, Bisantz & Drury (2000); McGrath et al. (2025)	FULL 12-ITEM SCALE: Distrust (reverse-scored, items 1–5): System is deceptive; Behaves underhanded; I am suspicious of intent/action/output; I am wary of system; Will have harmful outcome Trust (items 6–12): I am confident in system; Provides security; Has integrity; Is dependable; Is reliable; I can trust system; I am familiar with system VALIDATED SHORT FORM (S-TIAS, 3 items): Confident in system; System is reliable; I can trust system	3 ITEMS: TR1: I can rely on LLM's outputs as helpful starting point TR2: LLM will behave predictably TR3: Comfortable acting on LLM-assisted plan after I review it	Three trust items to minimise respondent burden while preserving coverage of confidence, predictability, and calibrated willingness to rely on LLM-assisted plans (including "appropriate use" qualifiers such as "as a helpful starting point" and "after I review it"). A validated three-item short form exists for the Trust in Automation Scale, supporting the use of brief trust batteries in AI contexts (McGrath et al., 2025).

				Boundary note: Because trust and PCA both touch reliability-related concerns, the PCA error-checking items (PCA8–PCA10) are treated as a priori contamination risks. In later waves, discriminant validity is assessed using HTMT (Henseler et al., 2015) alongside standard measurement-model checks. If trust–PCA overlap remains high, rewording should continue to emphasise user agency (e.g., "LLM support helps me catch errors…" rather than "the LLM catches errors…").
5. PCA (Perceived Cognitive Assistance)	NOVEL CONSTRUCT – This study	NO CANONICAL SCALE – first validation. Concept and item universe defined via reflexive thematic analysis of a three-tier corpus: 8 semi-structured interviews with active retail traders; 44 curated YouTube investor conversations; 24 qualitative and mixed-method studies on robo-advice, social/copy trading, and related advisory technologies (2015–2025), documented in Supplementary material C. Analysis yielded six recurring PCA themes: structural support	16 items (1–7 agreement scale; 1 = strongly disagree, 7 = strongly agree): PCA1: Using an LLM provides a structured path from my initial trading idea to a concrete order. PCA2: Using an LLM helps me break complex trades into manageable steps. PCA3: LLM support enables me to translate my market view into a precise, executable trade plan. PCA4: LLM support helps me structure multi-leg or conditional orders in a more organised way. PCA5: With LLM help, I can process more information at once without feeling overloaded. PCA6: LLM support helps me filter and isolate the specific information required for a trading decision. PCA7: LLM support helps me connect information from different sources into a coherent picture.	Coverage: structural scaffolding (PCA1–4); cognitive-load management and information triage (PCA5–7); error and failure-point checking (PCA8–10); scenario and volatility navigation (PCA11–13); metacognitive learning and exploration (PCA14–16). Plan: Use all 16 items in Wave 1 (PCA block plus joint EFA with TAM, Trust, self-efficacy) and retain 8–10

		(from idea to order); cognitive-load management; error-checking / quality control; volatility and scenario navigation; understanding why a strategy fits a view; exploration and autonomy-compatible expansion of the trading repertoire. The canonical PCA "content universe" for this project is defined as an over-specified 16-item pool covering all six themes.	PCA8: Using an LLM enhances my ability to spot inconsistencies or gaps in my trading plans. PCA9: LLM support helps me notice when numbers, dates, or assumptions in my plan do not line up. PCA10: LLM support prompts me to scrutinize my trade for potential failure points before I commit capital. PCA11: LLM support helps me compare alternative tactics for the same view side by side. PCA12: LLM support helps me think through what-if scenarios and plan for different market paths. PCA13: LLM support helps me keep my bearings and stay organised when markets move quickly. PCA14: LLM support helps me understand the rationale behind a strategy's suitability for my market view. PCA15: LLM support expands the range of trading scenarios I am able to mentally simulate. PCA16: LLM support facilitates reflection on my decision-making process during trade review.	items based on factor loadings, cross-loadings, and content coverage. In Wave 2, estimate a full measurement model (PCA, PU, PEOU, Trust, TPB, risk tolerance) with standard discriminant-validity checks (HTMT; Henseler et al., 2015; Fornell & Larcker, 1981) and incremental-validity tests for outcomes where PCA is theoretically expected to add beyond TAM and Trust.