DecisionCoach: A Multi-Agent Conversational System for Developing Decision-Making Competence Through Coaching Practice

Abstract. Developing transferable decision-making competence is difficult: learners often grasp decision concepts yet fail to execute high-quality processes in authentic contexts in higher education and training. This paper presents DecisionCoach (decisioncoach.io), an AI-mediated experiential learning platform that operationalizes decision-quality instruction through practice-based coaching simulations. Learners act as decision coaches to an LLM-driven student persona, eliciting information, clarifying objectives and tradeoffs, testing assumptions, and supporting commitment-to-action. We describe the system design, including modular orchestration for scenario and socio-emotional coherence, and an assessment layer that scores learner utterances using the Decision Quality Institute's six-element framework. DecisionCoach provides formative feedback and instructor-facing analytics for simulation–reflection–debrief integration. A pilot deployment demonstrates feasibility and measurable shifts from advice-giving toward structured inquiry.

Keywords: applied AI in education, simulation-based learning, decision coaching, decision quality, human computer interaction, large language models, learning through teaching

1 Introduction

1.1 The Decision-Making Education Challenge

Decision-making competence is among the most consequential skills an individual can develop. From career choices and financial planning to organizational strategy and personal relationships, the quality of our decisions shapes life outcomes across every domain [1,2]. Yet despite its importance, decision-making remains poorly taught in formal education. Students may learn domain knowledge—engineering, business, medicine—without developing systematic approaches to the decisions that knowledge must inform.

The challenge is fundamentally pedagogical: decision-making skill cannot be transmitted through lecture or textbook. Reading about good decisions does not produce good decision-makers, just as reading about swimming does not produce swimmers. Effective development requires practice with authentic decisions and feedback on decision process quality—conditions that traditional educational structures struggle to provide at scale.

Professional programs in fields like medicine, law, and business have long recognized this gap, developing case-based pedagogies that expose students to decision scenarios [3,4]. Yet case studies remain vicarious experiences; students analyze others' decisions rather than practicing their own decision processes. Simulation and role-play offer more active engagement but face authenticity and scalability constraints. The result is a persistent gap between knowing about good decisions and being able to make them.

1.2 The Learning-Through-Teaching Insight

Research in the learning sciences reveals a powerful but underutilized pathway: we learn most deeply when we teach others [5,6,7]. The act of explaining externalizes thinking, exposes gaps in understanding, and consolidates knowledge in ways that passive study cannot achieve [6,8]. Students who expect to teach material learn it more thoroughly than those studying for their own tests [7]. The "protégé effect" demonstrates that learners work harder—and learn more—when responsible for another's understanding [9].

This insight suggests a novel approach to decision-making education: rather than teaching people about decisions, have them coach others through decisions. The coaching process forces externalization of decision-making frameworks. To help another person frame a decision, one must understand framing. To guide someone toward meaningful alternatives, one must recognize when options are missing. To push a client toward reliable information, one must appreciate what good information looks like.

The core pedagogical mechanism is this: *by learning to coach, you learn to decide.* The skills developed in guiding others through decisions transfer directly to one's own decision-making. The Decision Quality framework becomes internalized not through memorization but through repeated application in coaching contexts.

1.3 From Insight to System: DecisionCoach

Translating this insight into scalable education requires solving the practice partner problem. Coaching practice traditionally requires willing clients with genuine decisions—a scarce resource in educational settings. Peer practice introduces self-consciousness and inauthenticity; classmates know their partner isn't really facing the presented dilemma. Large language models now enable a solution: AI personas can simulate clients facing authentic decisions with emotional complexity, ambivalence, and resistance that require sophisticated coaching responses.

DecisionCoach addresses this challenge through AI-mediated practice. The system presents learners with "Jamie," an LLM-based persona facing a career decision (switching majors). As learners attempt to coach Jamie through this decision, an automated assessment engine evaluates their coaching moves against the six elements of Decision Quality Science [10]. This dual architecture—realistic persona plus rubric-based assessment—provides the authentic practice and immediate feedback that decision-skill development requires.

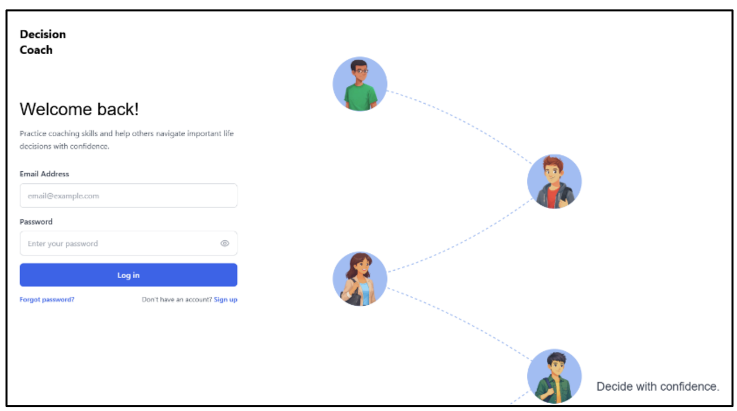

Figure 1Screenshot from the Decisioncoach.io homepage

Critically, DecisionCoach is not designed to train professional coaches—though it could serve that purpose. The system targets anyone seeking to improve their decision-making: students facing career choices, professionals navigating organizational decisions, individuals confronting major life transitions. By practicing the coaching role, these learners develop decision-making competence applicable to their own lives.

1.4 Contributions

This paper makes four contributions: (1) We present the design and implementation of DecisionCoach, an AI-mediated learning system that develops decision-making competence through coaching practice. (2) We describe the technical architecture including persona design, multi-agent orchestration, and rubric-based automated assessment. (3) We report baseline assessment results from an ongoing semester-long study (N=30), establishing pre-instruction performance levels and identifying systematic competency gaps that inform instructional design. (4) We discuss implications for integrating AI-mediated practice into decision-skill curricula and outline the study phases currently underway.

2 Background and Related Work

2.1 Decision Quality Science

Decision Quality (DQ) Science emerged from the decision analysis tradition pioneered by Howard [11], offering a framework for evaluating decisions independent of their outcomes. This distinction is foundational: conventional evaluation conflates decision quality with luck, crediting poor processes that happen to succeed while penalizing sound processes that encounter unfavorable results. DQ separates what decision-makers can control—the process—from what they cannot—the outcome—enabling meaningful assessment and skill development.

The framework identifies six elements that constitute a well-made decision [12,13]. Useful Framing establishes the decision's boundaries and success criteria; poor framing—asking the wrong question or excluding key stakeholders—undermines everything that follows. For Jamie, useful framing might reveal that the decision isn't simply "engineering versus design" but "how do I build a career that honors both family expectations and personal fulfillment?" Meaningful Alternatives ensures consideration of genuinely distinct options rather than false binaries; research on premature closure shows that decision-makers routinely truncate option generation, accepting inferior alternatives because better ones were never considered [14,15]. Reliable Information addresses what the decision-maker needs to know and how confident they can be—decisions made on assumptions or anecdotes carry unnecessary risk. Clear Values and Tradeoffs makes explicit what matters and how alternatives compare on valued dimensions, drawing on Keeney's value-focused thinking: understanding objectives and values should precede the generation of alternatives [16]. Sound Reasoning integrates information and values through logical analysis, ensuring conclusions follow from premises. Finally, Commitment to Action converts insight into implementation—identifying concrete next steps, assigning ownership, and establishing timelines. A decision without commitment is merely an intention.

The framework employs a "weakest link" principle: decision quality is limited by the weakest element, not averaged across them. A decision with excellent framing but no commitment remains incomplete. This principle emphasizes comprehensive coverage over isolated excellence and directly informs DecisionCoach's scoring approach.

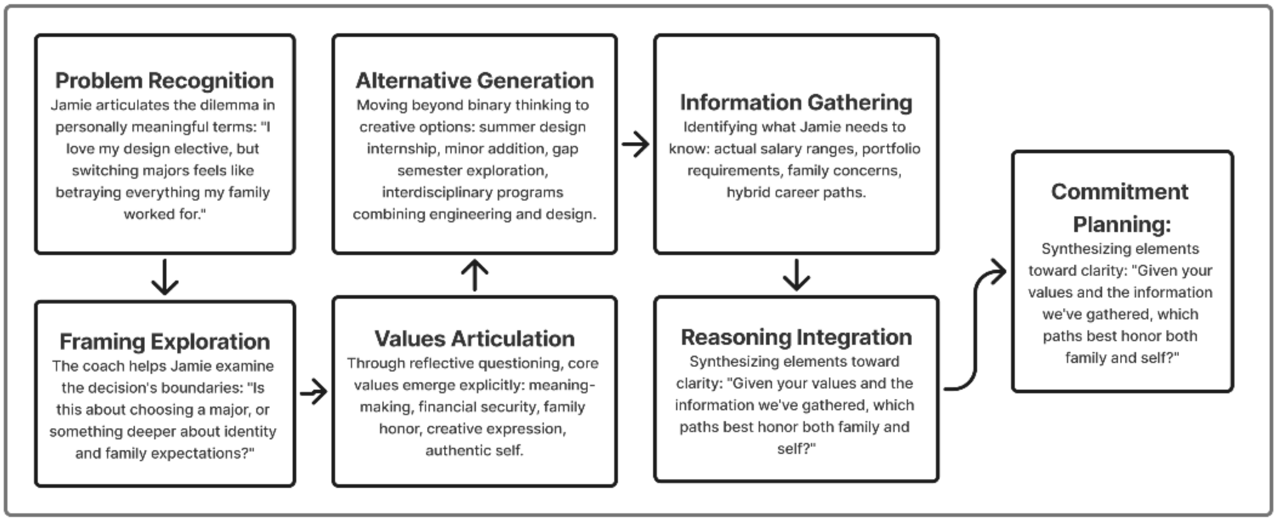

Figure 2Decision Quality Foundations 6 Elements of Decision Making

Several features make DQ suitable for educational contexts. The six elements are discrete and observable, enabling coaches to recognize and elicit each one. The framework is domain-independent, applicable to career decisions, organizational strategy, and personal choices alike. Most importantly, DQ focuses on process rather than outcome, enabling meaningful feedback even when the "right answer" is unknowable. Controlled experiments in professional education have demonstrated that DQ training can improve decision performance [17], validating the framework's teachability.

2.2 Learning Through Teaching

The pedagogical power of teaching has been documented across learning contexts. Chi et al. [8] demonstrated that self-explanation during learning produces deeper understanding than passive study. Roscoe and Chi [6] showed that tutors learn while tutoring, with the act of explanation consolidating and restructuring knowledge. Fiorella and Mayer [5] established the "learning by teaching" effect across multiple studies: expecting to teach material—and then teaching it—produces superior learning compared to studying for one's own test [7].

The "teachable agent" paradigm extends this insight to AI systems. Biswas et al. [18] introduced learning-by-teaching agents as an educational software paradigm, and Leelawong and Biswas [19] developed Betty's Brain, where students teach a computer agent by constructing concept maps; responsibility for the agent's learning can motivate deeper engagement. Chase et al. [9] found that teachable agents can increase learner effort and learning through the protégé effect: learners invest more when responsible for another's understanding.

DecisionCoach inverts this paradigm: rather than teaching an agent domain content, learners coach an agent through a decision process. The coaching role externalizes decision-making frameworks, revealing both the learner's understanding and their gaps. A learner who cannot help Jamie generate alternatives likely struggles with option generation in their own decisions.

2.3 Intelligent Tutoring and Conversational AI in Education

Intelligent Tutoring Systems (ITS) have demonstrated effectiveness approaching human tutoring [20], with meta-analytic reviews reporting meaningful learning gains relative to conventional instruction [21]. However, traditional ITS focus on domain knowledge transfer through problem-solving practice. Decision-making competence requires different pedagogical approaches: authentic scenarios, emotional complexity, and process feedback rather than right-answer feedback.

Conversational AI has emerged as a promising modality for educational applications, including LLM-based chatbots [22,23]. Unlike structured intelligent tutoring systems, conversational systems can sustain natural dialogue while embedding pedagogical scaffolds [22]. In mental health, early evaluations of automated conversational agents (e.g., Woebot) showed that users can engage meaningfully with an AI persona around emotional content [24]. Career counseling chatbots have also shown preliminary effectiveness for supporting career decision-making and information delivery [25], though many systems still emphasize content over decision process.

DecisionCoach occupies a distinct position: neither pure tutoring (teaching decision concepts) nor pure counseling (supporting decisions), but rather decision-skill development through coached practice. The conversational modality provides emotional realism while the assessment engine maintains focus on learnable coaching behaviors.

2.4 Gaps and Positioning

Despite advances in both decision education and educational AI, no existing system combines: (a) learning-through-teaching pedagogy, (b) emotionally realistic AI personas, (c) DQ-framework-aligned assessment, and (d) immediate formative feedback for decision-skill development. DecisionCoach addresses this gap by providing scalable practice infrastructure that gives students authentic repetitions and gives educators interpretable evidence of learning.

3 System Design

3.1 Design Goals

DecisionCoach was designed with four primary goals: (1) Authenticity—the simulated client must feel sufficiently real that learners engage as they would with an actual person facing a genuine decision. (2) Scalability—the system must support classroom deployment without requiring expert human evaluators for each session. (3) Interpretability—assessment must produce actionable feedback, not just scores, but insight into specific coaching behaviors and competency gaps. (4) Integration—the system must fit standard instructional patterns rather than requiring fundamental course redesign.

3.2 Persona Architecture: Jamie

The Jamie persona presents as a sophomore engineering major at a public university, carrying intersecting pressures: first-generation college attendance, family expectations of financial stability, and emerging passion for design. This configuration reflects documented patterns in major-switching decisions where students navigate between pragmatic security and personal fulfillment [14]. The persona was designed to be immediately recognizable to our target user group: early-stage college students facing similar decisions.

Jamie maintains consistency through structured backstory elements: grew up in a working-class family where college represents upward mobility; discovered unexpected creativity in an elective design course; feels torn between honoring family sacrifice and pursuing authentic interests. These details surface naturally in conversation rather than through exposition. For example, when discussing values, Jamie might say: "My parents worked double shifts to get me here… how can I tell them I want to study something they see as risky?"

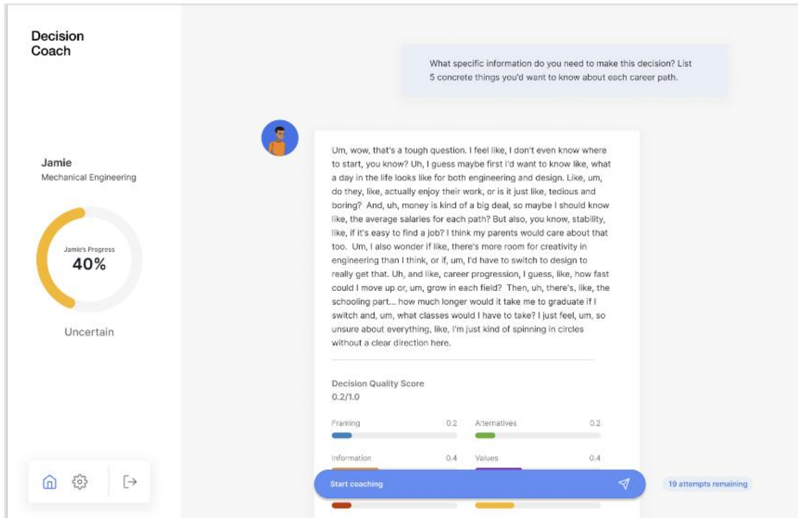

Figure 3 Decision Coach Chat Interface. The interface includes [1] 'Decision State' Progress Meter, [2] DQ Framework progress bars, [3] Turns Limit (bottom right.)

The persona includes an emotional progression model governing Jamie's trajectory from initial guardedness through reflection toward potential commitment. This progression responds to coaching quality: effective coaching moves Jamie toward clarity and action, while ineffective coaching (e.g., premature advice) triggers realistic resistance. Jamie may push back, express doubt, or redirect the conversation—behaviors that require coaches to demonstrate genuine DQ competence rather than following scripts.

3.3 Assessment Engine

The DecisionCoach assessment engine evaluates each coaching utterance against the six DQ elements using rubric-driven prompts. Scoring follows a 0.0–1.0 scale with behaviorally-anchored levels: 0.0–0.2 indicates the element is absent, counterproductive, or the persona is merely parroting suggestions without independent thinking; 0.3–0.4 reflects minimal presence with vague mention but no depth; 0.5–0.6 represents adequate presence with moderate effectiveness; 0.7–0.8 indicates strong presence with clear impact and independent analysis; and 0.9–1.0 reflects excellent execution with sophisticated integration.

The overall session score uses a "weakest link" principle—the minimum across all six dimensions—emphasizing comprehensive coverage rather than strength in isolated areas. This design choice reflects the DQ framework's emphasis on balance: a decision with excellent framing but no commitment remains incomplete. Assessment calibration was conducted through iterative comparison with expert DQ practitioners, with preliminary inter-rater reliability showing acceptable agreement ($r = 0.71$), though validation remains ongoing.

3.4 Technical Implementation

DecisionCoach is implemented as a lightweight web application with three primary components. The Conversational Interface is a React-based chat frontend providing turn-based interaction with session state tracking and progress visualization. The Persona Agent uses GPT-4o (OpenAI) to generate Jamie's responses based on persona prompts, emotional state, conversation history, and what DQ elements have or haven't been explored; prompting is intentionally conservative, with stable base prompts defining identity and constraints while short dynamic updates inject minimal state. The Assessment Agent is a separate GPT-4o instance that evaluates each coach utterance against the DQ rubric, generating per-turn scores, rationale, and dimension coverage status; this separation ensures assessment independence from response generation.

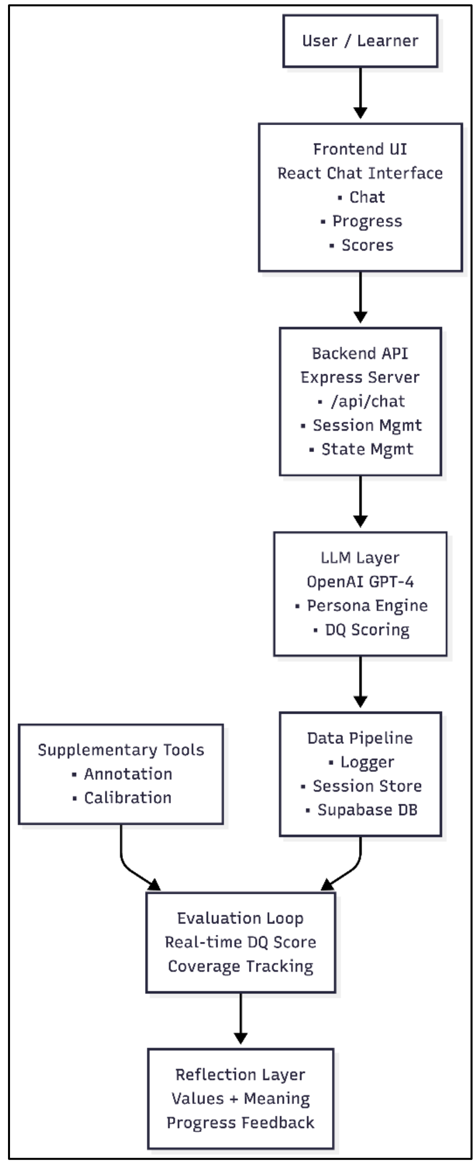

Figure 4 System Architecure

Sessions proceed through a turn-based cycle: coach utterance → assessment scoring → persona state update → Jamie response → display and logging. The system enforces a 10-turn limit, creating strategic constraint that transforms coaching from open-ended dialogue into purposeful practice. Context maintenance tracks conversation state and DQ coverage to support coherent, progressive coaching and prevent repetitive exploration of already-addressed elements.

3.5 Feedback Integration

DecisionCoach provides feedback at two levels. *Per-turn feedback* displays after each coach utterance: individual dimension scores (0.0–1.0), brief rationale explaining the scoring, and coverage status indicating which elements have reached threshold (≥0.3) versus remain unaddressed. *Session-level feedback* appears at session completion: aggregate scores across all six dimensions, coverage summary, target achievement comparison, and session summary including total turns and completion status. This dual-level feedback enables learners to adjust coaching dynamically during sessions while providing comprehensive assessment for reflection.

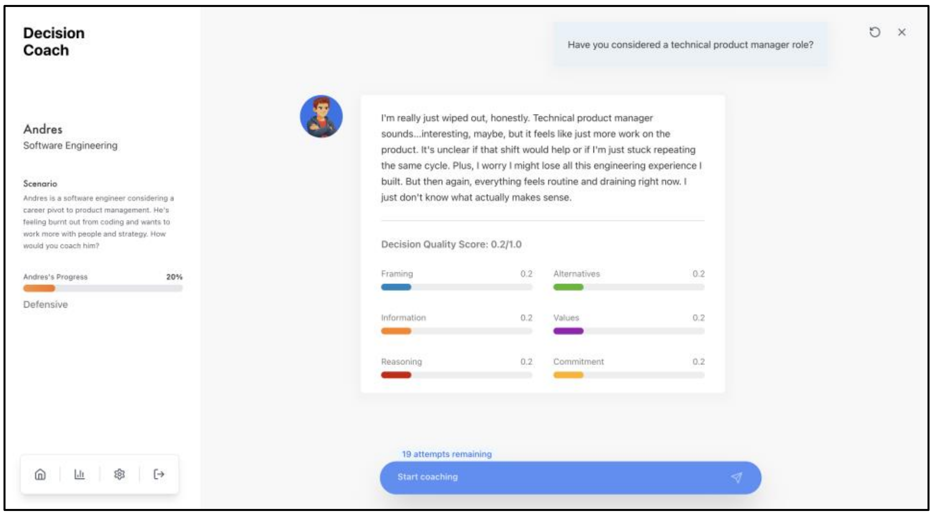

Figure 3 Screenshot of a chat interface shows the conversation flow with progress indicators.

3.6 Privacy and Ethical Considerations

Privacy and ethical considerations shaped design decisions throughout. The system explicitly presents itself as a simulation for learning purposes, avoiding any deception about AI nature. Data logging follows strict protocols: no personally identifiable information beyond session identifiers, explicit consent requirements, and transparent data use policies

Content guardrails prevent harmful directions: the system never dismisses family concerns as irrelevant, avoids privileged assumptions about financial flexibility, and maintains respect for all career paths. These ethical constraints operate invisibly, ensuring conversations remain supportive and inclusive across diverse student populations.

4 Instructional Integration

4.1 Course Integration Model

DecisionCoach is designed for integration into a semester-long decision-making curriculum following a pre-post instructional model. Students complete four phases: (1) baseline assessment with the Jamie persona before receiving DQ instruction, establishing individual starting points; (2) classroom instruction on DQ methods, coaching techniques, and the inquiry-before-advice principle; (3) deliberate practice with additional personas featuring diverse decision challenges and enhanced per-turn feedback; and (4) post-assessment using the original Jamie scenario to measure improvement.

A typical implementation includes a brief pre-brief on the six DQ elements and what "good coaching" looks like, including examples of inquiry versus advice; an individual DecisionCoach session with real-time element scores visible; a required reflection memo where students cite specific coaching turns and explain what they would do differently; and an in-class debrief where the instructor surfaces common patterns, discusses exemplary moves, and facilitates peer discussion of strategies.

Because the learner role is "coach," the activity avoids the awkwardness of asking students to disclose personal dilemmas publicly while still cultivating transferable decision habits. For instructors, the value lies not only in practice volume but in interpretable evidence. Element-level scores and transcripts support formative grading (participation plus reflection quality), targeted mini-lessons (e.g., how to generate high-quality alternatives), and repeat-practice assignments that demonstrate improvement over time.

4.2 Target Contexts

The system fits naturally in contexts where students face consequential choices: first-year seminars introducing decision-making frameworks, career development courses requiring students to articulate values and evaluate options, advising and peer-mentor training programs, leadership and ethics courses examining decision processes, and professional education programs in business, healthcare, and public policy. The flipped model—assigning DecisionCoach interactions as homework followed by group debriefs—maximizes scarce class time for collaborative learning while ensuring all students receive individualized practice.

5 Study Design and Baseline Assessment

5.1 Study Overview

The full study employs a pre-post instructional design spanning one academic semester. This paper reports results from Phase 1—the baseline assessment—which establishes pre-instruction performance levels and identifies competency gaps that inform the instructional phases now underway. Phase 2 provides explicit DQ instruction addressing the gaps identified at baseline. Phase 3 offers deliberate practice with new personas and enhanced feedback. Phase 4, at semester's end, reassesses students using the original Jamie scenario to measure instructional effectiveness through within-subject comparison.

5.2 Research Questions

The baseline assessment addressed three questions: RQ1—How do students perform when coaching through decisions prior to DQ instruction? RQ2—Do students show implicit learning within a single session, even without explicit framework training? RQ3—Which decision-making competencies are weakest at baseline, suggesting priority areas for instruction?

5.3 Participants

Thirty students enrolled in an undergraduate decision-making course completed baseline assessments during the first week of the semester, prior to any instruction on the Decision Quality framework. The sample included students from diverse academic backgrounds including engineering, design, business, and humanities. No prior training in decision-making frameworks, coaching methodologies, or Decision Quality was required or provided. Seven participants completed multiple sessions during the baseline phase, enabling preliminary analysis of cross-session patterns without instruction.

5.4 Procedure

Participants received minimal orientation to establish authentic baseline performance uncontaminated by framework knowledge. They were told only that they would practice helping a client (Jamie) work through a decision, with their approach evaluated. No instruction on DQ elements was provided—the assessment measured untrained intuitions about effective coaching. During the session, participants had 10 conversational turns to help Jamie make progress, with real-time per-turn feedback visible showing dimension scores but not explaining the framework. Participants knew the target was 70% DQ Overall score. Upon completion, session-level feedback displayed aggregate scores and threshold status.

5.5 Measures and Analysis

The baseline assessment generated 454 coach utterances across 30 sessions. Primary measures included per-utterance DQ scores (automated assessment across six dimensions), session-level aggregates (mean, minimum, and trajectory), within-session change (early turns 1–3 vs. late turns 8–10), element-level performance rankings, qualitative analysis of high and low performers, and message characteristics (length, question presence). Within-session improvement was tested using paired-samples t-tests comparing early versus late turn means. Effect sizes (Cohen's d) were calculated for significant differences.

6 Baseline Results

6.1 Pre-Instruction Performance (RQ1)

No participant achieved the 70% target threshold at baseline—an expected finding given that assessments occurred before DQ instruction. Session means ranged from 0.15 to 0.58 ($M = 0.27$, $SD = 0.09$). The distribution was concentrated at lower performance levels: 47% of sessions scored in the "Low" band (0.15–0.25), 40% in "Medium" (0.25–0.35), and only 14% in "High" or "Very High" bands (>0.35).

Table 1. Baseline Performance Distribution

Performance Band	DQ Score Range	Sessions	Percentage
Low	0.15–0.25	14	47%
Medium	0.25–0.35	12	40%

High	0.35–0.45	2	7%
Very High	>0.45	2	7%

This baseline distribution establishes the gap that instruction aims to close. The variance (range = 0.43) suggests the task differentiates skill levels even before training, providing a meaningful measure for detecting post-instruction improvement.

6.2 Within-Session Learning (RQ2)

Despite low absolute performance, participants showed significant improvement within sessions even without explicit instruction. Comparing early turns (1–3) to late turns (8–10), mean DQ Overall score improved from 0.20 to 0.32 ($\Delta = +0.12$, $t(29) = 4.87$, $p < .001$, $d = 0.89$). This represents a 60% improvement over baseline and suggests the feedback mechanism supports implicit learning.

Table 2. Within-Session Improvement by DQ Element

Measure	Early Turns	Late Turns	Δ
DQ Overall	0.20	0.32	+0.12**
Commitment to Action	0.17	0.34	+0.17**
Reliable Information	0.16	0.30	+0.14**
Sound Reasoning	0.21	0.33	+0.12**
Meaningful Alternatives	0.18	0.29	+0.11**
Useful Framing	0.22	0.32	+0.10*
Clear Values & Tradeoffs	0.26	0.35	+0.09*

Note: *$p < .05$, **$p < .01$

Improvement occurred across all six elements, with the largest gains in initially-weakest areas: Commitment to Action (+0.17, 100% improvement) and Reliable Information (+0.14, 88% improvement). This pattern—greatest improvement where starting performance was lowest—suggests the feedback mechanism effectively scaffolds coaches toward addressing neglected elements.

6.3 Competency Gaps Informing Instruction (RQ3)

Element-level performance revealed systematic gaps that directly inform the instructional phase now underway:

Table 3. Baseline DQ Element Performance

DQ Element	Mean	Instructional Implication

Clear Values & Tradeoffs	0.31	Relative strength; build on natural values engagement
Useful Framing	0.28	Moderate; teach reframing techniques
Sound Reasoning	0.28	Moderate; introduce structured comparison tools
Meaningful Alternatives	0.24	Weakness; teach option expansion techniques
Commitment to Action	0.24	Weakness; teach "sized action" planning
Reliable Information	0.23	Greatest weakness; prioritize information-gathering skills

Because coaching requires applying decision-making frameworks, these baseline weaknesses reflect gaps in students' own decision-making repertoires. Students who never pushed Jamie toward information-gathering likely neglect information-gathering in their own decisions. The instructional phase addresses these gaps directly, with particular emphasis on the three weakest elements.

6.4 Behavioral Patterns

Dominant failure mode. The most significant finding concerns coaching approach: only 20% of utterances contained questions. The dominant pattern was advice-giving without understanding—jumping to solutions before exploring the decision structure, alternatives, or information needs. This "premature solutioning" is itself a well-documented decision-making error, suggesting that coaching behavior mirrors personal decision-making tendencies. The instructional phase now emphasizes a "question quota"—requiring two or three clarifying questions before any recommendation.

Success pattern. The highest-performing session (M = 0.58, peak = 0.72) demonstrated qualitatively different behavior characterized by "structured empathy": combining emotional attunement with systematic decision exploration. This coach followed a recognizable arc—clarify the decision frame, expand options, weigh trade-offs, convert to actionable next steps—with each move building on previous understanding. This pattern is now taught explicitly as a coaching model.

Message characteristics. Message length correlated positively with quality scores ($r = 0.49$, $p < .001$). Short messages (<100 characters) averaged 0.23 DQ score; long messages (>200 characters) averaged 0.32. This relationship likely reflects engagement depth: sophisticated coaching requires elaboration, while brief responses indicate shallow engagement or premature closure.

Cross-session patterns. Seven participants completed multiple sessions during baseline. Their performance did not reliably improve across sessions without instruction—some improved, some declined, some remained stable. This finding reinforces the necessity of explicit instruction: practice alone, without framework training and structured reflection, does not produce durable skill transfer.

7 Discussion

7.1 Baseline Findings and Instructional Response

The baseline assessment establishes that untrained students perform well below the 70% target, with systematic weaknesses in information-gathering, alternative generation, and commitment-building. These findings directly shaped the instructional phase now underway. Instruction emphasizes question-first orientation (addressing the 80% advice-giving pattern), information elicitation techniques (addressing the weakest element at $M = 0.23$), and commitment scaffolds teaching "sized action"—identifying the smallest meaningful next step with owner and timeline.

The within-session improvement ($d = 0.89$) demonstrates that even without explicit instruction, feedback produces measurable learning. This suggests that combining explicit instruction with practice should yield substantially greater gains—a hypothesis the post-assessment will test. The cross-session finding—that practice without instruction does not produce reliable improvement—underscores that the instructional phase is necessary, not merely supplementary.

7.2 Design Lessons

Several design insights emerged from baseline analysis. *Strategic constraint through turn limits:* The 10-turn limit transformed coaching from meandering dialogue into strategic practice. Coaches learned to prioritize high-impact questions, resist premature problem-solving, and navigate the tension between depth and coverage. *Structured flexibility:* Pure open-ended conversation lost focus, while rigid scripts felt mechanical. The sweet spot involved flexible frameworks—conversational patterns that could adapt while ensuring coverage, like jazz improvisation with chord progressions rather than classical music with fixed notes. *Emotional before analytical:* Sessions that began with values and emotional exploration before analytical frameworks showed better engagement. Starting with decision trees or probability assessments triggered resistance. This sequencing—"heart before head"—honors how humans actually process complex choices. *Time-boxing as liberation:* The exploration semester concept emerged as particularly powerful. By creating contained experiments rather than permanent commitments, coaches could help Jamie act despite uncertainty. *Commitment as persistent weakness:* Across all sessions, commitment scores were consistently lowest (often 0.2–0.3 even in otherwise strong sessions). Participants found it easier to explore values and discuss information needs than to help create concrete, actionable plans. This suggests commitment-building requires different coaching skills than exploration—a gap the practice personas will specifically target.

7.3 Limitations

Several limitations constrain interpretation of baseline findings. *Sample composition:* The baseline included 30 students from a single course at one institution. While diverse in academic background, findings may not generalize to other populations. *Single scenario:* All baseline sessions used the Jamie persona; the practice personas will test whether skills transfer across decision types. *Automated scoring:* While preliminary calibration showed acceptable agreement ($r = 0.71$), comprehensive validation across diverse coaching styles remains ongoing. *Pre-post design limitations:* Without a control group receiving no instruction, we cannot fully isolate instructional effects from maturation or repeated testing; however, the cross-session baseline data suggesting no improvement without instruction partially addresses this concern.

8 Ongoing Development

8.1 Study Phases in Progress

Phase 2 (DQ Instruction) is currently underway, providing explicit training on the six DQ elements with particular emphasis on the three weakest areas identified at baseline: information-gathering, alternative generation, and commitment-building. Instruction includes coaching technique demonstrations, the "question quota" requirement, and structured reflection on baseline transcripts.

Phase 3 (Deliberate Practice) introduces additional personas beyond Jamie, each presenting distinct decision challenges. Planned personas include returning adult students balancing education with family responsibilities, international students navigating visa and cultural constraints, and students facing financial pressure requiring tradeoff clarity. Each persona preserves emotional depth and consequential stakes while surfacing different decision dynamics. Practice sessions feature enhanced per-turn feedback that explicitly names DQ elements present or missing, enabling targeted skill development.

Phase 4 (Post-Assessment) will occur at semester's end, with students completing the original Jamie scenario under conditions identical to baseline. Comparing pre and post scores on the same scenario enables within-subject measurement of instructional effectiveness and identification of which competency gaps instruction successfully addressed.

8.2 Technical Development

Technical improvements address specific weaknesses observed at baseline. *Enhanced reasoning support:* Structured analytical tools—decision matrices, trade-off prompts, scenario analysis—will be embedded within conversational flow, offered contextually when reasoning scores stall. *Improved alternative generation:* Creative thinking prompts including constraint removal exercises ("If you had unlimited resources..."), analogical reasoning ("What would someone in a similar situation do?"), and hybrid solution generation will trigger when coaches converge prematurely on binary options. *Stronger commitment scaffolding:* Specific prompts for timeline creation, contingency planning, and accountability structures will address the persistent commitment weakness, helping coaches move from insight facilitation to action planning.

8.3 Methodological Extensions

Beyond process metrics, we are developing outcome measures including creativity of alternatives (non-binary options, hybrid solutions), values-alignment of choices, and robustness to uncertainty. Longitudinal tracking will assess whether coaching skills transfer to students' own decision-making and whether improvements persist beyond the course. The evaluation rubric itself—operationalizing abstract concepts like "good judgment" into observable behaviors—represents a contribution applicable beyond this system, potentially useful for evaluating human advisors, comparing coaching approaches, or training practitioners.

9 Conclusion

This paper introduced DecisionCoach, an AI-mediated system for developing decision-making competence through coaching practice. The core insight—that coaching others through decisions builds one's own decision-making capability—offers a scalable approach to an educational challenge that traditional methods address incompletely.

Baseline assessment established that untrained students perform well below target, with systematic weaknesses in information-gathering, alternative generation, and commitment-building. Within-session improvement ($d = 0.89$) demonstrates that feedback produces learning even without explicit instruction—suggesting that combining instruction with practice should yield substantially greater gains. The dominant failure pattern—advice-giving without inquiry—reveals where untrained decision-makers struggle and directly informs the instructional phase now underway.

The contribution extends beyond any specific educational context. Career-deciding students, strategy-setting professionals, life-navigating individuals—all face decisions that their education poorly prepared them to make. DecisionCoach demonstrates that practicing the coaching role, externalized through AI-mediated dialogue, can develop the decision-making competence that formal education too often neglects.

As large language models continue advancing, opportunities for learning-through-teaching applications will expand. DecisionCoach represents a step toward realizing this potential for decision-making education—a domain where skill development matters profoundly but scalable practice has remained elusive. Post-assessment results, to be reported following semester completion, will provide evidence of whether the instructional model achieves its aims.

Acknowledgments

This work was supported in part by the Decision Education Foundation, the Office of Professional and Continuing Education and the Decision Quality Center at Illinois Institute of Technology.

References

[1] R. H. Thaler and C. R. Sunstein, Nudge: Improving Decisions About Health, Wealth, and Happiness. New Haven, CT: Yale University Press, 2008.
[2] D. Kahneman, Thinking, Fast and Slow. New York: Farrar, Straus and Giroux, 2011.
[3] D. A. Garvin, "Making the case: Professional education for the world of practice," Harvard Magazine, vol. 106, no. 1, pp. 56-65, 2003.
[4] C. C. Lundberg, P. Rainsford, J. P. Shay, and C. A. Young, "Case writing reconsidered," Journal of Management Education, vol. 25, no. 4, pp. 450-463, 2001.
[5] L. Fiorella and R. E. Mayer, "The relative benefits of learning by teaching and teaching expectancy," Contemporary Educational Psychology, vol. 38, no. 4, pp. 281-288, 2013.
[6] R. D. Roscoe and M. T. H. Chi, "Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions," Review of Educational Research, vol. 77, no. 4, pp. 534-574, 2007.

[7] J. F. Nestojko, D. C. Bui, N. Kornell, and E. L. Bjork, "Expecting to teach enhances learning and organization of knowledge in free recall of text passages," Memory & Cognition, vol. 42, no. 7, pp. 1038-1048, 2014.
[8] M. T. H. Chi, M. Bassok, M. W. Lewis, P. Reimann, and R. Glaser, "Self-explanations: How students study and use examples in learning to solve problems," Cognitive Science, vol. 13, no. 2, pp. 145-182, 1989.
[9] C. C. Chase, D. B. Chin, M. A. Oppezzo, and D. L. Schwartz, "Teachable agents and the protégé effect: Increasing the effort towards learning," Journal of Science Education and Technology, vol. 18, no. 4, pp. 334-352, 2009.
[10] R. A. Howard and A. E. Abbas, Foundations of Decision Analysis. New York: Pearson, 2015.
[11] R. A. Howard, "Decision analysis: Practice and promise," Management Science, vol. 34, no. 6, pp. 679-695, 1988.
[12] C. Spetzler, H. Winter, and J. Meyer, Decision Quality: Value Creation from Better Business Decisions. Hoboken, NJ: Wiley, 2016.
[13] Strategic Decisions Group, "Decision Quality: The fundamentals," Decision Quality Institute, 2018. [Online]. Available: https://www.decisionquality.com
[14] J. E. Russo and P. J. H. Schoemaker, Decision Traps: Ten Barriers to Brilliant Decision-Making and How to Overcome Them. New York: Simon & Schuster, 1989.
[15] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," Science, vol. 185, no. 4157, pp. 1124-1131, 1974.
[16] R. L. Keeney, Value-Focused Thinking: A Path to Creative Decision Making. Cambridge, MA: Harvard University Press, 1992.
[17] H. Winter, C. Spetzler, and J. Meyer, "Teaching decision quality in professional schools: Results from a controlled experiment," Decision Analysis, vol. 18, no. 3, pp. 189-208, 2021.
[18] G. Biswas, K. Leelawong, D. Schwartz, N. Vye, and The Teachable Agents Group at Vanderbilt, "Learning by teaching: A new agent paradigm for educational software," Applied Artificial Intelligence, vol. 19, no. 3-4, pp. 363-392, 2005.
[19] K. Leelawong and G. Biswas, "Designing learning by teaching agents: The Betty's Brain system," International Journal of Artificial Intelligence in Education, vol. 18, no. 3, pp. 181-208, 2008.
[20] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," Educational Psychologist, vol. 46, no. 4, pp. 197-221, 2011.
[21] J. Kulik and J. Fletcher, "Effectiveness of intelligent tutoring systems: A meta-analytic review," Review of Educational Research, vol. 86, no. 1, pp. 42-78, 2016.
[22] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, and G. Kasneci, "ChatGPT for good? On opportunities and challenges of large language models for education," Learning and Individual Differences, vol. 103, p. 102274, 2023.
[23] S. Wollny, J. Schneider, D. Di Mitri, J. Weidlich, M. Rittberger, and H. Drachsler, "Are we there yet? A systematic literature review on chatbots in education," Frontiers in Artificial Intelligence, vol. 4, p. 654924, 2021.
[24] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial," JMIR Mental Health, vol. 4, no. 2, e19, 2017.

[25] H. Han, B. Park, and K. Seo, "A Self-Determination Theory-based Career Counseling Chatbot: Motivational Interactions to Address Career Decision-Making Difficulties and Enhance Engagement," in Extended Abstracts of the 2025 CHI Conference on Human Factors in Computing Systems (CHI EA '25), Association for Computing Machinery, 2025, Art. 48.

Appendix

Element	Basic (0.2–0.4)	Proficient (0.5–0.7)	Advanced (0.8–1.0)
Framing	Mentions decision exists	Explores boundaries and stakeholders	Reframes problem constructively
Values	Lists surface preferences	Articulates priorities with rationale	Integrates values with identity
Alternatives	Binary options only	Multiple distinct paths generated	Creative synthesis of options
Information	Basic facts mentioned	Relevant evidence identified	Critical gaps and sources specified
Reasoning	Simple trade-off noted	Systematic comparison attempted	Uncertainty explicitly integrated
Commitment	Vague future intentions	Specific next steps identified	Accountability structures established

Rubric for DQ Assessment Engine