Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 100–109
https://doi.org/10.54941/ahfe1008245

A Behavioral Observation Tool for Diagnosing Human–AI Interaction Profiles in Engineering Design Education

Stefano Filippi

University of Udine - DPIA Dept., Udine, Italy

ABSTRACT

The integration of generative Artificial Intelligence into engineering design education has created a practical challenge for instructors: how to recognize, in real time, how individual students are interacting with AI systems during laboratory sessions. While empirical research has identified three recurrent Human–AI interaction profiles—Cognitive Adapters, Functional Executors, and Skeptical Observers—existing diagnostic approaches rely on survey instruments and post-hoc analysis that are unsuitable for classroom use. This paper proposes a Behavioral Observation Tool (BOT) designed for use by instructors during laboratory sessions. The BOT consists of 17 behavioral indicators organized across five dimensions—goal formulation, output interpretation, iterative behavior, agency distribution, and reflective engagement—rated on a three-point scale. A weighted scoring system produces normalized profile scores that classify students into dominant, mixed, or indeterminate profiles. Profile-specific instructional guidelines, derived from an adaptive Human–AI interaction framework, translate classification outcomes into concrete teaching responses. An Excel-based scoring instrument accompanies the tool, automating calculation and classification without requiring statistical expertise. The BOT bridges the gap between empirical interaction research and everyday instructional practice, supporting reflective, differentiated, and human-centered AI integration in engineering education.

Keywords: Human–AI interaction, Engineering education, Behavioral observation, Interaction profiles, Instructional design, Generative AI

INTRODUCTION

The growing integration of generative Artificial Intelligence (AI) into engineering education is reshaping how students engage with design tasks, particularly in activities involving product representation and conceptual modeling. As AI tools become standard components of laboratory sessions, they introduce new forms of interaction that go beyond simple tool use: students must formulate intentions as prompts, interpret machine-generated outputs, iterate on representations, and continuously negotiate control between their own reasoning and system-generated suggestions. This shift has significant implications for learning, cognition, and the development of design competence (Amershi et al., 2019; Shneiderman, 2022).

Recent research has identified three recurrent Human–AI interaction profiles among engineering students: Cognitive Adapters, who engage AI

Received June 8, 2026; Revised October 1, 2026; Accepted October 7, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

reflectively and iteratively; Functional Executors, who rely on AI primarily for efficient task completion; and Skeptical Observers, who exhibit low trust and limit AI integration (Filippi, in press-b). These profiles have been further developed into an adaptive interaction framework linking each configuration to differentiated forms of instructional support, grounded in Norman's action cycle and Schön's reflection-in-action (Filippi, in press-a).

However, the identification of these profiles has so far relied on self-report surveys and post-hoc data analysis—methods incompatible with the real-time demands of laboratory teaching. This paper addresses this gap by proposing a Behavioral Observation Tool (BOT) for diagnosing Human–AI interaction profiles during laboratory sessions. The BOT consists of 17 behavioral indicators across five observational dimensions, rated on a three-point scale. A weighted scoring system produces normalized profile scores enabling dominant, mixed, or indeterminate classification. An Excel-based scoring instrument accompanies the tool, automating calculation and classification without requiring statistical expertise.

BACKGROUND

Interaction Profiles in AI-Mediated Design Tasks

The integration of AI tools into engineering education has generated substantial research on students' attitudes, trust, and perceived usefulness toward these systems (Kasneci et al., 2023; Zawacki-Richter et al., 2019). Recent work has identified three recurrent Human–AI interaction profiles among engineering students engaged in product representation tasks (Filippi, in press-b). The profiles were derived through cluster analysis applied to four composite indicators—trust calibration, evaluation of AI contributions, conceptual depth of interaction, and cognitive load—and validated through structured laboratory observations coded according to Norman's action cycle and Schön's reflection-in-action (Norman, 1984; Schön, 1983).

Cognitive Adapters exhibit high trust calibration and conceptual engagement. They iterate extensively, refine goals through prompt reformulation, critically interrogate AI outputs, and engage in sustained reflection-in-action. Functional Executors achieve effective task outcomes through streamlined interaction, accepting satisfactory outputs with limited evaluative iteration and delegating a portion of agency to the AI system. Skeptical Observers display the lowest levels of trust and conceptual engagement, encountering breakdowns at multiple stages of the action cycle and frequently reverting to manual representational strategies.

These profiles are not fixed user categories but recurrent interaction tendencies that can vary across tasks and contexts. Their recurrence across a multi-year cohort of undergraduate and graduate engineering students suggests they capture meaningful and stable aspects of student engagement with AI during design activities.

The Adaptive Interaction Framework

The profiles have been further developed into an adaptive interaction framework linking each configuration to specific forms of instructional support (Filippi, in press-a), organized around four dimensions: guidance in

goal formulation, interpretability of outputs, support for iterative exploration, and distribution of agency. Cognitive Adapters benefit from low-to-moderate guidance, high iteration support, and balanced agency. Functional Executors benefit from moderate guidance and an AI-leaning agency configuration. Skeptical Observers benefit from high guidance, high interpretability support, and a human-leaning agency configuration. This framework provides the theoretical foundation for the instructional recommendations embedded in the BOT.

The Gap: From Research Instruments to Classroom Practice

Despite the empirical solidity of the interaction profiles and the practical relevance of the adaptive framework, no instrument currently exists that allows an instructor to diagnose interaction profiles through direct, in-session observation and act on that diagnosis immediately. Existing approaches require research expertise, post-hoc data processing, and conditions incompatible with real-time laboratory teaching. This gap is pedagogically consequential: instructors who cannot identify how individual students interact with AI cannot provide timely, differentiated support, potentially reinforcing suboptimal engagement patterns. The BOT proposed in this paper is designed to close this gap.

TOOL DEVELOPMENT

Development Rationale

The BOT was developed following a top-down logic grounded in the empirical and theoretical work described in the Background section. The development process consisted of three sequential steps: (1) identifying the behavioral correlates of each interaction profile from existing observational evidence; (2) selecting and operationalizing behavioral indicators that are simultaneously theoretically grounded, observationally accessible, and diagnostically informative; and (3) constructing a scoring system that maps observed behaviors onto interaction profiles in a transparent and practically usable way. The BOT is therefore not an independent instrument but a structured translation of existing empirical findings into an operational classroom tool. Its validity derives from the empirical and theoretical foundations on which it rests, rather than from independent psychometric validation—a limitation acknowledged in the Discussion section.

Indicator Selection

Behavioral indicators were derived from the observational protocol used in the original profile identification study (Filippi, in press-b), grounded in Norman's action cycle and Schön's reflection-in-action (Norman, 1984; Schön, 1983). Five observational dimensions were identified as theoretically meaningful and practically accessible: goal formulation, output interpretation, iterative behavior, agency distribution, and reflective engagement.

Within each dimension, indicators were selected according to three criteria: observability—the behavior must be directly visible without requiring inference; discriminability—the behavior must differ systematically across profiles; and coverage—the set must represent the range of behavioral variation within the dimension. Three indicators per dimension were selected initially, yielding 15 items. Two additional indicators were added to ensure adequate coverage of the Functional Executor profile, bringing the final set to 17 items.

Scoring System

Each indicator contributes to one or more profile scores through a weighted mapping: weight 2 for strong and exclusive profile indicators, weight 1 for shared or weaker indicators. Raw scores are normalized to a percentage of the theoretical maximum for each profile:

$$Score(profile) = [\Sigma\ (weight \times observed\ rating)] / [theoretical\ maximum] \times 100$$

Theoretical maxima are: Adapter = 32, Executor = 18, Observer = 18. Classification follows three tiers: dominant profile (one profile exceeds both others by ≥ 15 percentage points); mixed profile (two profiles within 15 points of each other, with guidance provided for the three most common hybrid configurations: Adapter/Executor, Executor/Observer, and Adapter/Observer); and indeterminate (all three within 15 points, recommending a second observation session).

THE OBSERVATION TOOL

Structure and Format

The BOT is designed as a single-page observation form completable during or immediately after a laboratory session. For each of the 17 behavioral indicators, the instructor assigns a rating of 0 (not observed), 1 (occasionally observed), or 2 (frequently observed). Ratings should reflect the overall pattern of behavior rather than isolated occurrences. Table 1 presents the unified BOT form, incorporating behavioral indicators, profile weights, and rating columns.

Table 1: Behavioral Observation Tool (BOT): indicators, profile weights, and rating form.

ID	Dimension	Behavioral Indicator	Adpt	Exec	Obs	0	1	2
A1	Goal Formulation	Rewrites prompt multiple times before accepting output	2			☐	☐	☐
A4	Goal Formulation	Avoids long or detailed prompts			2	☐	☐	☐
A5	Goal Formulation	Compares output with own expectations before proceeding	2			☐	☐	☐
B1	Output Interpretation	Examines output at length before acting	1		1	☐	☐	☐
B2	Output Interpretation	Accepts output immediately and moves to next task		2		☐	☐	☐

(Continued)

Table 1: Continued.

ID	Dimension	Behavioral Indicator	Adpt	Exec	Obs	0	1	2
B3	Output Interpretation	Compares multiple alternative outputs	2			☐	☐	☐
B5	Output Interpretation	Ignores parts of output not understood		1	1	☐	☐	☐
C1	Iterative Behavior	Generates multiple variants of the same concept	2			☐	☐	☐
C2	Iterative Behavior	Stops at the first acceptable solution		2		☐	☐	☐
C3	Iterative Behavior	Goes back to reformulate after seeing result	2			☐	☐	☐
C5	Iterative Behavior	Alternates AI use with manual work			2	☐	☐	☐
D2	Agency Distribution	Uses AI output as-is without modification		2		☐	☐	☐
D3	Agency Distribution	Abandons AI and reverts to traditional method			2	☐	☐	☐
D5	Agency Distribution	Works autonomously without seeking validation	1	1		☐	☐	☐
E2	Reflective Engagement	Works silently without externalizing process		1	1	☐	☐	☐
E3	Reflective Engagement	Spontaneously stops to compare result with task requirements	2			☐	☐	☐
E5	Reflective Engagement	Takes notes or annotates observations on output	2			☐	☐	☐
		Theoretical maximum	32	18	18			

Note. Adpt = Cognitive Adapter; Exec = Functional Executor; Obs = Skeptical Observer.
Weights: 2 = strong profile indicator; 1 = shared or weak indicator; blank = no contribution. Scores: 0 = Not observed; 1 = Occasionally; 2 = Frequently. Theoretical maximum scores reflect the sum of maximum weighted ratings (2) across all contributing indicators.

Scoring Procedure

Once all 17 indicators have been rated, the instructor computes the score for each profile in three steps:

- **Step 1—Compute raw profile scores.** $RawScore(profile) = \Sigma\ (weight \times observed\ rating)$
- **Step 2—Normalize to percentage.** $Score(profile) = RawScore(profile) / TheoreticalMaximum(profile) \times 100$ where theoretical maxima are: Adapter = 32, Executor = 18, Observer = 18.
- **Step 3—Apply classification rules** as defined in the Scoring System subsection.

The Excel Scoring Tool

To eliminate calculation burden, the BOT is accompanied by an Excel-based scoring instrument. The instructor enters the 17 ratings into a single input sheet; the tool automatically computes raw scores, normalized percentages,

and profile classification, displaying the result with a visual indicator. No statistical expertise is required. The tool extends a spreadsheet-based methodology previously applied in engineering education research (Filippi et al., in press-c) to real-time instructional support.

Worked Example

Table 2 illustrates the scoring procedure with a hypothetical observation record.

Table 2: Worked example: observation record and scoring.

ID	Rating	Adapter (w×r)	Executor (w×r)	Observer (w×r)
A1	2	2 × 2 = 4		
A4	0			2 × 0 = 0
A5	2	2 × 2 = 4		
B1	2	1 × 2 = 2		1 × 2 = 2
B2	0		2 × 0 = 0	
B3	2	2 × 2 = 4		
B5	0		1 × 0 = 0	1 × 0 = 0
C1	2	2 × 2 = 4		
C2	0		2 × 0 = 0	
C3	2	2 × 2 = 4		
C5	0			2 × 0 = 0
D2	0		2 × 0 = 0	
D3	0			2 × 0 = 0
D5	2	1 × 2 = 2	1 × 2 = 2	
E2	0	1 × 0 = 0	1 × 0 = 0	
E3	2	2 × 2 = 4		
E5	2	2 × 2 = 4		
	Raw score	32	2	2
	Normalized score	100%	11%	11%

Classification: Dominant Adapter—the Adapter score exceeds both others by more than 15 percentage points. This student would be classified as a Cognitive Adapter, and the instructor would respond by reducing guidance and increasing exploratory freedom as described in the Application Guidelines section.

APPLICATION GUIDELINES

When to Observe

The BOT is designed for use during laboratory sessions in which students interact with AI tools to perform product representation or conceptual design tasks. To obtain a reliable observation, the session should last at least 45–60 minutes—sufficient to cover at least two full interaction cycles—and involve a sufficiently open-ended task that allows behavioral differences to emerge naturally. A single observation session is sufficient when a dominant profile

is obtained; a second session is recommended for mixed or indeterminate classifications. A focused observation window of 20–30 minutes, covering at least one complete interaction cycle from goal formulation through output evaluation, is generally sufficient to rate most indicators. Indicators that remain unobserved should be rated 0, as their absence is itself diagnostically informative for certain profiles.

How to Record

The BOT is completed by a single observer—typically the course instructor or a teaching assistant—who circulates among students while maintaining a non-intrusive presence. Brief notes during the session support accurate rating at the end. Instructors should be aware that some indicators may be less visible in fully digital environments—for example, E5 (takes notes or annotates) should be rated based on any observable analog behavior, or rated 0 if absent.

What to Do With the Result

Profile classification informs real-time adjustments to the form and intensity of instructional support, derived directly from the adaptive interaction framework (Filippi, in press-a).

Cognitive Adapters engage reflectively and iterate extensively. The primary risk is over-guidance: unsolicited suggestions can interrupt productive exploration. Instructors should adopt a light-touch approach—asking open questions rather than providing answers, and allowing extended iteration time. If convergence difficulty arises, a soft deadline or comparison criterion is preferable to a direct solution.

Functional Executors achieve outcomes efficiently but with limited reflective depth. The primary risk is reinforcement of superficial engagement. Instructors should actively promote comparison and iteration—requesting at least one alternative before finalizing, or introducing a variation constraint that requires revisiting the AI output.

Skeptical Observers exhibit low trust and limited AI engagement. The primary risk is disengagement. Instructors should provide structured support in prompt formulation—offering templates or guided reformulation exercises—and help interpret AI outputs by discussing their content, assumptions, and evaluation criteria.

Handling Mixed and Indeterminate Profiles

For mixed profiles, the instructional response combines elements from the two relevant configurations: Adapter/Executor students need iteration prompts when efficiency-oriented behavior becomes dominant; Executor/Observer students need support in both output interpretation and reflective comparison; Adapter/Observer students need their critical stance validated while being encouraged toward structured AI engagement. For indeterminate classifications, a neutral approach is recommended until a second observation clarifies the profile.

Integration Into Course Design

Beyond individual sessions, the BOT can serve as a formative assessment instrument across a course. A baseline observation at the start of semester informs overall instructional strategy; repeated observations at mid-course and end-of-course document whether interaction patterns evolve as students gain AI literacy (Filippi et al., in press-c), complementing traditional performance-based assessment (Kasneci et al., 2023; Shneiderman, 2022).

DISCUSSION

Contributions

This paper has proposed a Behavioral Observation Tool for diagnosing Human–AI interaction profiles in engineering design education. The BOT represents a novel contribution at the intersection of Human–AI interaction research and instructional practice, addressing a gap that existing approaches have left open: the absence of a practical, real-time instrument that instructors can use to recognize and respond to students' interaction patterns during laboratory sessions. The BOT is grounded in empirically validated interaction profiles and a theoretically coherent adaptive framework, ensuring that its indicators and scoring system reflect meaningful behavioral differences. The accompanying Excel scoring tool eliminates calculation burden and supports immediate deployment, extending a methodology previously applied in research-oriented frameworks to real-time instructional support (Filippi et al., in press-c). Together, these features position the BOT as a bridge between the descriptive and prescriptive dimensions of Human–AI interaction research.

Theoretical and Practical Implications

From a theoretical perspective, the BOT demonstrates that empirically identified profiles can be translated into observable behavioral indicators without loss of theoretical coherence. The five observational dimensions correspond directly to the interaction dynamics identified in the adaptive framework (Filippi, in press-a) and to the stages of Norman's action cycle and Schön's reflection-in-action that underpin it (Norman, 1984; Schön, 1983). The agency distribution dimension in particular operationalizes a construct—how control and initiative are shared between student and system—that has been identified as central to interaction quality (Shneiderman, 2022; Amershi et al., 2019) but rarely made visible at the classroom level.

From a practical perspective, the BOT supports a shift in how AI integration is conceptualized in engineering curricula: from attending to whether students use AI, to attending to how they engage with it. This aligns with calls for interaction-centered pedagogies that prioritize reflective engagement and critical evaluation over task efficiency (Kasneci et al., 2023; Tholander & Jonsson, 2023).

Limitations and Future Directions

Several limitations should be acknowledged. The BOT has not yet been subjected to formal psychometric validation—inter-rater reliability, test-retest stability, and convergent validity with survey-based instruments remain to be

established. The 15-point classification threshold was set on rational grounds rather than through empirical optimization. The observational dimensions were developed in the context of product representation tasks at a single European university, and their transferability to other disciplines and task types requires further investigation.

Future work should pursue field validation through concurrent survey and observation studies, longitudinal application across course cycles to document profile evolution (Filippi et al., in press-c), and digital implementations—such as tablet-based observation apps integrated with learning analytics platforms—that extend the BOT's reach beyond individual sessions.

CONCLUSION

This paper has proposed a Behavioral Observation Tool (BOT) for diagnosing Human–AI interaction profiles in engineering design education. Grounded in empirically validated interaction profiles and an adaptive interaction framework developed in prior work, the BOT translates research-level knowledge into a practical, instructor-oriented instrument deployable during laboratory sessions without requiring survey administration, data processing, or statistical expertise.

The tool consists of 17 behavioral indicators organized across five observational dimensions—goal formulation, output interpretation, iterative behavior, agency distribution, and reflective engagement—rated on a three-point scale. A weighted scoring system and normalized classification procedure assign students to one of three interaction profiles—Cognitive Adapter, Functional Executor, or Skeptical Observer—or to a mixed or indeterminate configuration when behavioral evidence is insufficient. Profile-specific instructional guidelines translate classification outcomes into concrete teaching actions implementable within the same session.

By making interaction dynamics visible and actionable in real time, the BOT supports a shift toward interaction-centered pedagogies in which instructors attend not only to what students produce but to how they engage with AI in the process of producing it—consistent with broader calls for human-centered, reflective, and differentiated approaches to AI integration in higher education (Shneiderman, 2022; Amershi et al., 2019; Kasneci et al., 2023).

REFERENCES

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S.T., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human–AI Interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3290605.3300233

Filippi, S. (in press-a). Adaptive Human–AI Interaction Strategies in Engineering Education: A Profile-Based Framework for AI-mediated Product Representation. Education and Information Technologies. Springer.

Filippi, S. (in press-b). Human–AI Interaction in Product Representation: Cognitive and Educational Implications in Engineering Design Courses. Lecture Notes in Computer Science. Springer.

Filippi, S., Vaglio, E., & Motyl, B. (in press-c). Integrating Generative AI in Engineering Education: A Longitudinal Framework on Personality, Roles, and Perceptions. AHFE Conference Proceedings. https://doi.org/10.54941/AHFE-Paper-0045

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Norman, D.A. (1984). Stages and Levels in Human–Machine Interaction. International Journal of Man-Machine Studies, 21(3), 365–375. https://doi.org/10.1016/S0020-7373(84)80054-1

Schön, D.A. (1983). The Reflective Practitioner: How Professionals Think in Action. Basic Books, New York.

Shneiderman, B. (2022). Human-Centered AI. Oxford University Press.

Tholander, J., & Jonsson, M. (2023). Design Ideation with AI: Sketching, Thinking and Talking with Generative Machine Learning Models. Proceedings of the 2023 ACM Designing Interactive Systems Conference, 1930–1940. https://doi.org/10.1145/3563657.3596014

Zawacki-Richter, O., Marín, V.I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16, 39. https://doi.org/10.1186/s41239-019-0171-0