AI-SUPPORTED SOCRATIC TUTORING CHATBOT IN MANUFACTURING EDUCATION: A THEORY-INFORMED CDIO IMPLEMENTATION

Abboy Verkuilen

Technology, Innovation & Society, Mechanical Engineering, The Hague University of Applied Sciences

ABSTRACT

This paper describes theory-informed AI-powered chatbots as Socratic tutors in a first-year Manufacturing Technology course (Productietechniek 1) at The Hague University of Applied Sciences. The chatbots combine large language models (ChatGPT and Gemini) with Retrieval-Augmented Generation (RAG), drawing on curriculum-specific resources including lecture materials and worked examples. Unlike much AI-in-education research, the implementation is grounded in established learning theories: Cognitive Load Theory, Explicit Instruction, Self-Regulated Learning, Formative Assessment, Self-Determination Theory, and Sociocultural Learning Theory. Students used the chatbot collaboratively in small groups to analyse products and design production processes. Across five weeks of structured teacher observation in four parallel class groups, the two observers noted patterns consistent with the theoretical predictions: sustained on-task behaviour, more explicit technical reasoning in group dialogue, engagement with AI fallibility as a learning moment, and emerging articulation of what a professional product requires. These observations are offered as candidate design hypotheses rather than evaluated outcomes. In the subset of chatbot exchanges directly observed, incorrect AI output was frequently traceable to incomplete or imprecise student input, and these moments appeared to function productively within group discussion. This work contributes to the CDIO initiative by demonstrating how AI can enhance formative feedback, student agency, and professional competence development as a theory-informed conversational learning partner.

KEYWORDS

AI-supported learning, Socratic tutoring, formative assessment, manufacturing education, Cognitive Load Theory, Self-Regulated Learning, Standards: 2, 7, 8, 11

INTRODUCTION

The rapid development of generative AI tools such as ChatGPT and Gemini presents opportunities and risks for engineering education. Recent meta-analyses demonstrate positive effects of AI chatbots on student learning outcomes (Wu & Yu, 2024; Wang & Fan, 2025). However, the same research highlights significant risks: AI systems may produce inaccurate information through hallucination, encourage superficial learning through over-reliance, or undermine student agency when used as substitutes for thinking rather than supports for it (Danyaro et al., 2025; Hon, 2025).

The educational research literature reveals a concerning pattern: many studies of AI-based learning environments report student satisfaction and perceived learning benefits, but lack explicit grounding in established learning theories or critical discussion of pedagogical design (Ofosu-Ampong, 2024; Memarian & Doleck, 2024). This theoretical gap limits the capacity to understand why certain AI implementations succeed while others fail, and hampers the transfer of design principles across educational contexts.

Within the CDIO framework, AI introduces a clear educational challenge and opportunity. The CDIO Syllabus 3.0 emphasises that engineering graduates must develop capacities for critical thinking, professional reasoning, and collaborative problem-solving (Malmqvist et al., 2022). Professional engineering practice increasingly involves working with AI-driven tools across design, analysis, and manufacturing domains. Engineering programmes therefore bear responsibility to help students critically and responsibly engage with AI, while simultaneously strengthening key CDIO outcomes including reflective practice, feedback-rich learning, professional judgement, and collaborative reasoning.

This paper reports on AI-supported Socratic tutoring in the first-year course Productietechniek 1 (Manufacturing Technology 1) in the Mechanical Engineering programme at The Hague University of Applied Sciences. Approximately 150 students enrol annually, comprising diverse entrants from MBO, HAVO, VWO, and university transfer routes. This heterogeneity creates substantial variation in prior knowledge and learning approaches, which the AI tutoring system was designed to accommodate.

The course introduces fundamental manufacturing processes and requires students to design justified production process plans for real-world components. Two key design intentions guided the intervention. First, AI should function as a formative feedback instrument rather than a replacement for student reasoning. Second, AI use should strengthen peer learning, agency, and reflection rather than isolate students from collaborative learning dynamics.

The guiding research question is: *How can AI chatbots be deployed in manufacturing education such that they add didactic and pedagogical value while stimulating peer interaction and professional learning?* The paper presents the theoretical foundations, instructional design, technical configuration, and the learning dynamics the two observers noticed during this first implementation cycle, contributing to the growing body of CDIO scholarship on technology-enhanced active learning.

THEORETICAL FRAMEWORK

Although AI-in-education research has expanded rapidly since 2023, integration of established learning theories remains limited. Systematic reviews consistently identify this weakness (Batista et al., 2024; Liang et al., 2025). This implementation explicitly embeds chatbot design within six complementary learning theories, responding to CDIO community calls for evidence-based innovation (Edström et al., 2020).

Cognitive Load Theory

Cognitive Load Theory (CLT) explains how instructional design affects learning (Sweller, 1988, 1994). The theory distinguishes intrinsic load (element interactivity), extraneous load (suboptimal design), and germane load (schema construction). Effective instruction should manage intrinsic load through sequencing, minimise extraneous load, and promote germane load (Sweller et al., 1998, 2019). The AI tutoring system supports CLT-aligned design by providing short instructional turns, using structured questioning that reduces ambiguity, breaking complex tasks into smaller steps, and encouraging visual and procedural reasoning before process selection. Research on intelligent tutoring systems shows effect sizes of approximately d = 0.76 compared to conventional instruction (VanLehn, 2011; Ma et al., 2014; Kulik & Fletcher, 2016).

Explicit Instruction

Explicit Instruction is characterized by clear learning objectives, systematic presentation, modelling of cognitive processes, guided practice with feedback, and gradual release of responsibility (Rosenshine & Stevens, 1986; Rosenshine, 2012; Archer & Hughes, 2011). The approach benefits novice learners lacking schema structures for discovery-based instruction (Hughes et al., 2017). The chatbot incorporates these elements by articulating learning goals, modelling expert reasoning, offering guided practice through worked examples, and checking understanding through targeted questions. Students construct reasoning pathways themselves rather than receiving direct solutions—balancing explicit guidance with productive struggle.

Formative Assessment

Formative assessment powerfully influences student learning. Sadler (1989) established that effective formative assessment requires students to understand learning goals, perceive performance gaps, and take action to close them. Black and Wiliam's (1998) landmark review demonstrated effect sizes of 0.4 to 0.7 standard deviations, while Hattie (2009) found d = 0.73 for feedback interventions. The chatbot operationalizes these principles by eliciting current understanding before offering guidance, identifying misconceptions through probing, providing feedback through hinge questions rather than declarative judgement, and periodically revisiting learning goals.

Self-Regulated Learning

Self-Regulated Learning (SRL) theory describes how learners manage their cognitive processes, motivation, and behaviour (Zimmerman, 1989, 2000). Zimmerman's model identifies three cyclical phases: forethought, performance, and self-reflection. Pintrich and De Groot (1990) demonstrated that SRL components predict academic performance beyond cognitive ability alone, while Panadero (2017) highlighted the need for interventions that explicitly develop self-regulatory skills. The chatbot stimulates SRL by requiring students to plan before receiving guidance, monitor their understanding, adjust their prompts based on responses, and evaluate AI-generated proposals critically. Students observe a direct relationship between their input quality and AI output quality, reinforcing metacognitive control and countering over-reliance risks.

Self-Determination Theory

Self-Determination Theory (SDT) identifies three psychological needs: autonomy, competence, and relatedness (Deci & Ryan, 1985, 2000; Ryan & Deci, 2000). When supported, these needs foster autonomous motivation characterized by interest and integrated self-regulation. Autonomy-supportive practices predict greater engagement and deeper learning (Ryan, 2023). The chatbot addresses all three pillars. Autonomy is supported by student control of questioning pace and direction. Competence is supported through mastery-focused feedback rather than comparative evaluation. Relatedness is maintained through collaborative use in small groups.

Sociocultural Learning Theory

Sociocultural learning theory emphasizes that learning is a social process mediated by cultural tools and more capable others (Vygotsky, 1978). Wood, Bruner, and Ross (1976) operationalized Vygotsky's Zone of Proximal Development through their analysis of scaffolding. Lave and Wenger (1991) extended this through situated learning and legitimate peripheral participation. Because students use the chatbot collaboratively in small groups, learning becomes dialogic and socially mediated. The chatbot functions as a neutral third conversational partner that stimulates negotiation of meaning, joint decision-making, and justification of design choices.

Theoretical Integration

By integrating CLT, Explicit Instruction, Formative Assessment, SRL, SDT, and Sociocultural Theory, this paper responds to a significant gap in AI-based educational research where theoretical grounding is often limited or implicit. Each theory contributes distinct insights: CLT addresses cognitive architecture, Explicit Instruction provides sequencing principles, Formative Assessment guides feedback design, SRL emphasises metacognitive development, SDT addresses motivation, and Sociocultural Theory frames the social learning context. This theoretical triangulation provides a robust foundation for both design decisions and interpretation of observed effects.

This work builds on prior CDIO community emphasis on active learning (Edström & Kolmos, 2014), formative assessment (Hansen & Sindre, 2023), and student agency (Crawley et al., 2014), extending these themes into AI-supported learning. The framework aligns with CDIO Standards 8, 11, and 2.

COURSE CONTEXT AND CDIO ALIGNMENT

The scientific calculator reshaped engineering education in the 1970s not by giving students answers, but by shifting the locus of cognitive effort: away from manual computation and toward problem formulation, estimation, and interpretation (Ruthven, 2013). GenAI in the manufacturing classroom warrants the same framing. It is a cognitive tool embedded within the learning activity — not a substitute for instruction, and not a peripheral technology encountered only outside class. This framing matters because it determines how the tool should be aligned with educational standards. The Mechanical Engineering programme at The Hague University of Applied Sciences prepares students for applied professional engineering practice in the Dutch and international manufacturing industry, and follows CDIO principles with particular emphasis on the integration of disciplinary knowledge with professional skills development throughout the curriculum. The remainder of this section situates the course within that CDIO framework and, in the final subsection, maps the intervention explicitly to the CDIO Standards.

Course Description and Learning Outcomes

Productietechniek 1 (Manufacturing Technology 1) introduces fundamental manufacturing processes: forming (casting, forging, sheet metal forming), separating (machining, cutting), joining (welding, adhesive bonding, mechanical fastening), and relationships between material properties, process selection, and product quality. Learning outcomes align with CDIO Syllabus 3.0 competencies in disciplinary knowledge (Section 1), engineering reasoning and problem solving (Section 2.1), and conceiving and designing engineering processes (Section 4.3). Students are expected to analyze products to identify manufacturing considerations, select and justify processes based on technical and economic criteria, develop logically sequenced production process plans, and evaluate choices against quality, cost, and sustainability criteria.

Student Population Characteristics

The student population is heterogeneous, comprising entrants from MBO (practical technical experience), HAVO (limited technical background), VWO (strong theoretical but limited practical experience), and university transfer routes. This creates instructional challenges: substantially different prior knowledge, varied learning approaches, limited individual feedback opportunities in traditional instruction, and the need to bridge theoretical and practical knowledge. The AI tutoring system was designed specifically to address this heterogeneity by providing adaptive, individualized support within a structured learning framework.

Professional Assessment Structure

Professional assessment requires students to produce a production process plan demonstrating ability to apply manufacturing knowledge to realistic engineering problems. Students analyze a component's geometry, material, function, and quality requirements; select and justify manufacturing processes with explicit reasoning; sequence operations coherently; and evaluate choices against professional criteria. The AI chatbots support analysis, reasoning, and reflection during the formative learning phase—not for summative assessment. This distinction is important: the chatbot functions as a practice and feedback tool in a low-stakes environment, encouraging experimentation and learning from errors.

Alignment with CDIO Standards

Situating the chatbot-supported activity explicitly within the CDIO Standards (CDIO Initiative, 2020; Malmqvist et al., 2019, 2020) clarifies how the classroom pedagogy aligns with the framework and, equally importantly, how it responds to external signals from industry and society.

Standard 2 (Learning Outcomes).

The learning outcomes of Productietechniek 1 are deliberately aligned not only with CDIO Syllabus 3.0 Section 1 (disciplinary fundamentals) and Section 4.3 (conceiving and designing engineering processes), but also with Section 2.1 (engineering reasoning and problem solving), Section 2.4 (attitudes, thought, and learning), Section 2.5 (ethics, equity, and other responsibilities), and Sections 3.1–3.2 (teamwork and communication), precisely the competencies that external stakeholders now flag as decisive for employability. The World Economic Forum's Future of Jobs Report 2025 identifies AI and big data as the fastest-growing skill categories worldwide, with 86% of surveyed employers expecting AI and data analytics to drive business transformation and 39% of core skills projected to change by 2030 (World Economic Forum, 2025). In the Dutch manufacturing context specifically, FME, Smart Industry and Koninklijke Metaalunie's joint market research reports that 83% of Dutch SME manufacturers lack insight into which parts of their production process are suitable for AI-supported automation, and the FME Smart Industry Productiviteitsagenda 2026–2028 names AI literacy and workplace innovation as preconditions for restoring Dutch manufacturing productivity (FME, 2025). The Sociaal-Economische Raad (SER), in its advice “AI en werk” to the Dutch government, argues that the workforce needs critical thinking, digital skills, and domain-specific problem-solving capacity to work alongside AI systems responsibly rather than defer to them uncritically (SER, 2025), echoing earlier Rathenau Instituut framing of responsible AI adoption (Rathenau Instituut, 2023). The EU AI Act, effective February 2025, reinforces this through its AI literacy obligation for organisations developing or deploying AI. Read together, these industry and societal voices align closely with CDIO Syllabus 3.0 Section 2.4 and 2.5: graduates of a Dutch mechanical engineering programme must be able not only to apply manufacturing knowledge, but to reason critically about AI-generated proposals, judge their validity against disciplinary criteria, and work ethically with AI tools in a team setting. The learning outcomes of this course should be updated to reflect this industry rationale, and the chatbot-supported activity is the primary mechanism through which these updated outcomes are enacted in class.

Standard 7 (Integrated Learning Experiences).

Standard 7 requires that disciplinary knowledge be learned alongside professional skills in integrated activities that mirror engineering practice. The chatbot-supported design task directly mirrors a growing category of contemporary engineering work: a production engineer at a Dutch manufacturer using GenAI to triage design-for-manufacturing (DFM) issues, iteratively evaluating both process choices (which manufacturing route, at what cost and quality) and design choices (whether to modify geometry, tolerances, or material selection to improve manufacturability), with the AI tool as a first-pass conversational partner whose suggestions must then be critically verified against standards, supplier data, and the engineer's own judgement. Students in Productietechniek 1 are not simulating this workflow at a distance; they are performing a legitimate peripheral version of it (Lave & Wenger, 1991), integrating material science, process knowledge, professional reasoning, communication, and responsible AI use within a single activity.

Standard 8 (Active Learning).

The intervention operationalises Standard 8 through four concrete mechanisms observable in the classroom: (i) chatbot questions trigger peer discussion within the small group before a response is composed, making student-student dialogue a structural feature rather than an optional extra; (ii) prompt construction becomes a joint analytical task, in which students negotiate what to tell the chatbot and why; (iii) every chatbot output is treated as a proposal to be evaluated, not an answer to be copied, shifting cognitive work toward critique and justification; and (iv) the lecturer circulates between groups to deepen reasoning at moments where the chatbot alone would under-challenge the students. Active learning in this design is therefore produced by the interaction between chatbot, peers, and lecturer, not by any single agent.

Standard 11 (Learning Assessment).

Standard 11 is operationalised through a clear separation between formative and summative functions. The chatbot is used formatively only: students use it during class to test reasoning, identify gaps, and rehearse justifications. Formative feedback is sourced primarily from the lecturer and from peers within the group, with the chatbot playing a secondary role by prompting students to self-assess or to check their group's shared understanding. Because the chatbot's knowledge base includes worked examples of professional production process plans from previous cohorts, students who become stuck can request and interrogate concrete exemplars of the target artefact, consistent with Explicit Instruction principles on the role of worked examples for novice learners (Hughes et al., 2017). The chatbot does not generate the summative deliverable, but it makes the standard of a high-quality plan more visible during the learning phase. The summative assessment, the production process plan, is completed independently of the chatbot and assessed against a professional rubric by the lecturer. This separation preserves assessment validity while allowing the chatbot to extend formative feedback capacity beyond what a single lecturer can provide to approximately 150 students individually.

The intervention aligns with multiple CDIO Standards (CDIO Initiative, 2020; Malmqvist et al., 2019, 2020). Standard 2 (Learning Outcomes) is addressed through focus on engineering reasoning, problem-solving, and system thinking competencies, requiring students to integrate knowledge across materials, processes, and quality systems. Standard 7 (Integrated Learning Experiences) is supported as the activity integrates disciplinary knowledge with professional skills, mirroring workplace practices where engineers increasingly use AI tools. Standard 8 (Active Learning) is central to the design. The Socratic questioning approach requires students to engage deeply rather than passively receive information. Standard 11 (Learning Assessment) is operationalised through continuous formative assessment embedded in chatbot interactions, providing immediate feedback enabling real-time adjustment of understanding.

METHODOLOGY

Research Design

This study employs a qualitative, practice-based case study design within a design-based research (DBR) tradition (Anderson & Shattuck, 2012; McKenney & Reeves, 2018), appropriate for investigating pedagogical innovations in their authentic classroom context. DBR is well-suited to the iterative development of theory-informed educational interventions, where the researcher is simultaneously designer, practitioner, and investigator, and where the goal is to generate both usable knowledge for practice and transferable design principles for the wider community (Bakker, 2018). The case is bounded by a single course, Productietechniek 1, taught in the Mechanical Engineering programme at The Hague University of Applied Sciences during the 2025-2026 academic year, with approximately 150 first-year students distributed across parallel class groups. Within this case, the unit of analysis is the small student group (2–3 students) engaging collaboratively with the AI chatbot during in-class design sessions. The primary data source is structured teacher-researcher observation of these group interactions, supplemented by a secondary analysis of anonymised chatbot conversation logs. The study does not employ survey or interview methods; claims in this paper are therefore grounded in observational and log data only, with formal survey-based evaluation and inter-cohort comparison planned for a subsequent iteration of the design cycle. The author served as both lecturer and researcher, which offered privileged access to classroom dynamics but introduces a risk of observer bias that is explicitly addressed in the Limitations section. The study was conducted as part of routine educational practice; no identifying student data are reported, and we had no access to chatbot log excerpts, except for the numbers of chats conducted by the student groups.

Observation Method

Classroom observation was conducted jointly by the lecturer-researcher (the author) and a second lecturer teaching parallel class groups, providing two independent observer perspectives on the same intervention. Observations took place across five consecutive weekly sessions, distributed over four parallel class groups, yielding a total of approximately twenty observed session-hours. During each session, the two observers focused on two readily accessible layers of classroom interaction: (1) student–student dialogue within the small groups, which was easily observable at a distance and captured peer negotiation of meaning, disagreement, and joint decision-making; and (2) student–chatbot dialogue, observed opportunistically when students raised a question, requested lecturer input, or were prompted by the lecturer to share and justify their current chatbot exchange. Direct observation of chatbot dialogue was therefore partial rather than systematic, and biased toward moments where students themselves surfaced the interaction. Data were recorded in two complementary forms: brief contemporaneous field notes taken during the sessions, and more reflective post-session memos written by each observer shortly after class, capturing salient patterns, notable student utterances, and emerging questions. The analytic approach was deliberately light-touch and interpretive rather than formally systematic: observations were not coded against a predefined framework, and no thematic analysis in the Braun and Clarke (2006) sense was undertaken. Instead, the two observers compared field notes and memos informally across the five-week period, identifying recurrent patterns and moments that "stood out" with respect to engagement, reasoning quality, and student handling of AI fallibility. This pragmatic approach is consistent with first-cycle design-based research, where the aim is to surface candidate design hypotheses for subsequent, more rigorous evaluation (McKenney & Reeves, 2018), but it clearly limits the strength of inferential claims, as discussed in the Limitations section.

Chatbot conversation logs at the message level were not available to the research team; the learning platform recorded only session-level metadata. Across the five-week intervention, approximately 300 chatbot sessions were initiated by student groups. Each session typically contained multiple message exchanges, but the number of exchanges per session was not captured and varied substantially by group. The figure of 300 should therefore be read as an indicator of adoption intensity rather than as a measure of dialogue depth.

Because no chat content, audio, or video was recorded, and no personally identifiable student data were collected or analysed, the THUAS ethics committee confirmed that formal ethical review was not required for this study. The observations were conducted as part of the author's regular teaching practice, and all observed examples reported in this paper are paraphrased rather than quoted verbatim to further protect student anonymity.

TECHNICAL AND INSTRUCTIONAL DESIGN

The technical and instructional design reflects the theoretical framework above, with decisions driven by learning theory rather than technological capabilities alone (Labadze, Grigolia & Machaidze, 2023; Liu et al., 2025).

Retrieval-Augmented Generation Architecture

Each chatbot uses a commercial large language model (ChatGPT-4 or Gemini Pro) connected to a course-specific knowledge base through Retrieval-Augmented Generation (RAG). RAG architectures allow language models to ground responses in authoritative source materials while maintaining conversational flexibility (Liu et al., 2025). The knowledge base includes lecture PowerPoints based on Basisboek Productietechniek (the course textbook), structured guidance documents for production process design, worked example professional products from previous cohorts, and assessment formats, templates, and rubric criteria. This design limits irrelevant information and maintains curricular alignment. The system retrieves relevant content and incorporates it into responses, reducing hallucination likelihood while ensuring guidance aligns with process selection approaches taught in the course.

Didactic Prompt Design and Behavioral Rules

The chatbot's pedagogical behavior is shaped through carefully engineered instructions. Explicit behavioral instructions affect learning outcomes (Zhang, Z., et al., 2024). The chatbot is framed as a Socratic tutor that guides through questioning rather than telling, a formative assessor that identifies gaps and provides developmental feedback, and a metacognitive coach that prompts reflection on learning processes.

Core behavioral rules include the following. The chatbot should scaffold understanding through progressive questioning rather than providing full answers immediately. It should begin by clarifying student intent and current knowledge level. It should ask probing questions that reveal thinking and identify misconceptions. Explanations should be short and structured, addressing specific gaps rather than comprehensive lectures. It should adapt language level and technical depth to student responses, providing more scaffolding for struggling students and more challenge for advanced students.

Additional rules address AI-specific considerations. The chatbot warns students that hallucinations may occur and responses should be verified against course materials. It requires students to justify their reasoning rather than accepting suggestions uncritically, and promotes reflection on both the task and the learning process. Research on Socratic chatbots suggests this questioning approach enhances critical thinking and deeper learning compared to information-delivery approaches (Favero et al., 2024; Zhang, L., et al (2024)).

Learning Task Structure and Sequencing

The learning task follows a structured sequence developing systematic product analysis and process reasoning skills. Students first analyze their product in terms of material (type, properties, processing behavior), geometry (dimensions, tolerances, surface requirements), functional requirements (load bearing, wear resistance, corrosion resistance), context of use (environment, maintenance, lifecycle), and expected production quantities (prototype, batch, mass production). This structure reflects professional manufacturing process selection, ensuring students consider relevant factors before jumping to process choices. Students upload product images with their initial analysis, grounding conversation in concrete artefacts.

Following product analysis, the chatbot guides students through iterative design reasoning. It identifies missing information by questioning gaps, requests additional analysis where students have overlooked factors, proposes provisional process routes with reasoning, and asks students to critique and refine decisions by identifying potential problems or alternatives. This cycle continues until students develop a complete, justified production process plan. The chatbot's role is to guide and question rather than provide definitive answers, maintaining student ownership of design decisions.

Collaborative Learning Implementation

Students work with the chatbot in small groups of 2-3, sharing a single session. This design was driven by sociocultural learning theory and SDT considerations: collaborative use maintains social learning dynamics, supports relatedness needs that might be undermined by isolated AI interaction, enables peer evaluation of AI responses before acting on them, and distributes cognitive load across multiple students. The collaborative structure transforms the chatbot into a shared resource stimulating group discussion. When the chatbot asks a question or makes a suggestion, students discuss before responding, creating opportunities for peer learning and mutual explanation. Across implementation, students engaged in over 300 chatbot interactions. The collaborative structure also safeguards against over-reliance: multiple students evaluating responses reduces uncritical acceptance of incorrect information.

LIMITATIONS

Several limitations should be acknowledged, spanning both the evidence base and the methodological choices made in this first design-based research cycle. Framing these explicitly against established qualitative case study methodology (Creswell & Poth, 2018; Yin, 2018) helps clarify what this study can and cannot legitimately claim.

Evidence base

First, the absence of completed survey-based evaluation and inter-cohort comparison of assessment outcomes limits the evidence to observational and session-level log data. While observations provide valuable qualitative insights, they cannot establish causal relationships or quantify learning effects. A validated student survey and comparative analysis of assessment outcomes is planned for the next iteration. Second, the absence of message-level chatbot logs meant that dialogue depth, prompt quality evolution over time, and the precise point at which student groups shifted from answer-seeking to reasoning-seeking could not be quantified. The figure of approximately 300 chatbot sessions therefore indicates adoption intensity rather than dialogue depth, and reviewer-relevant questions about the number of exchanges required for critical evaluation to emerge cannot be answered from the current data.

Observer bias.

The dual role of the author as teacher, designer, and researcher introduces a substantial and unavoidable risk of observer bias (Yin, 2018). Both observers — the author and a second lecturer teaching parallel class groups — were personally invested in the pedagogical intervention and in the broader project of responsible AI integration at THUAS. Neither was blind to the design intent, the theoretical framework, or the desired learning behaviours. It is therefore plausible that the observers' attention was disproportionately drawn to confirmatory evidence (for example, instances of productive Socratic dialogue or critical evaluation of AI output) and less attuned to disconfirming evidence (for example, groups that used the chatbot superficially or disengaged). The absence of inter-rater reliability checks between the two observers, and the interpretive rather than formally coded analytic approach, compounds this limitation. Future cycles of the research will mitigate this through independent observation by a researcher not involved in course delivery, structured observation protocols, and formal consensus coding of field notes.

Partial observation of chatbot dialogue.

As noted in the Observation Method section, direct observation of student–chatbot exchanges was opportunistic rather than systematic: the observers captured dialogue primarily when students surfaced it through questions to the lecturer or when the lecturer prompted them to share their exchange. Student–student discussion was readily observable at a distance, but the content of chatbot conversations within groups that worked autonomously throughout a session was not accessible to the observers. Claims about the nature of student–chatbot dialogue in this paper are therefore based on a non-random subset of interactions and should be read as illustrative rather than representative.

Hawthorne and novelty effects.

Although students were not formally aware of being observed as research subjects (no chats, audio, or video were recorded, and formal ethical review was not required), they were explicitly aware that the lecturers were interested in how the new AI chatbot supported their learning. This creates conditions for both a Hawthorne effect — behavioural change in response to the perceived interest of authority figures — and a novelty effect associated with the introduction of an unfamiliar tool. While the sustained engagement observed across five weeks argues against pure novelty as the sole driver, neither effect can be ruled out on the basis of observational data alone. Longitudinal comparison across multiple cohorts, once the intervention is no longer novel, will be necessary to disentangle these confounds.

Single-site generalisability and transferability.

The study is situated in a single course at a single Dutch university of applied sciences, which constrains statistical generalisability in the conventional sense (Yin, 2018). However, following Creswell and Poth's (2018) notion of transferability in qualitative inquiry, the theoretical foundations that guided this design — Cognitive Load Theory, Explicit Instruction, Formative Assessment, Self-Regulated Learning, Self-Determination Theory, and Sociocultural Learning Theory — are broadly supported across disciplines and contexts in the educational research literature. The classroom observations reported here are consistent with the predictions of this literature, lending cautious support to the expectation that the underlying design principles would transfer to other engineering education contexts. What will vary substantially, however, is the implementation: the specific curricular knowledge base, the nature of the professional artefact students design, the balance of lecturer facilitation and chatbot scaffolding, and the disciplinary norms governing what counts as a justified design decision. The contribution of this paper is therefore best read as a set of transferable design principles rather than a replicable classroom recipe.

Long-term transfer.

Finally, the study does not track whether the observed learning effects persist beyond the course or transfer to professional engineering practice, where AI-mediated reasoning is increasingly expected. Ultimate educational value therefore remains partially unknown and will require longitudinal follow-up with graduates, which is beyond the scope of a first-cycle design-based study.

Taken together, these limitations are consistent with the exploratory, practice-oriented nature of first-cycle design-based research (McKenney & Reeves, 2018) and with Yin's (2018) view that single-case studies generate analytic rather than statistical generalisation. The findings presented here are offered as candidate design hypotheses to be tested, refined, and challenged in subsequent cycles and in replication by the wider CDIO community.

OBSERVED PATTERNS AND CANDIDATE DESIGN HYPOTHESES

The two observers identified recurrent patterns across the five-week period, surfaced through informal comparison of post-session memos rather than through formal coding. These patterns are reported here as candidate design hypotheses to be tested in subsequent DBR cycles through survey, inter-cohort comparison, and message-level log analysis. They are not offered as evidence that the design caused particular learning outcomes, and language of increase, improvement, or effectiveness should be read accordingly.

Engagement and On-Task Behaviour

Both observers recorded an impression of sustained engagement with the manufacturing process design activity across the five-week period. Because no systematic comparison against prior cohorts was undertaken, this impression should be read as hypothesis-generating rather than comparative. Novelty plausibly contributed to initial engagement. Whether the observed sustained attention reflects the pedagogical design, the collaborative format, the chatbot specifically, or residual novelty cannot be determined from the present data; this is a question for the planned inter-cohort comparison. The collaborative structure amplified engagement. The observers noted animated within-group discussion and debate about process choices, consistent with the sociocultural rationale for collaborative chatbot use. The comparison to 'individual work typically observed' is impressionistic; no matched comparison was performed. The chatbot functioned as a conversation starter and maintained momentum when group discussion flagged. These observations align with meta-analytic findings on AI chatbots and engagement (Wang & Fan, 2025) and SDT predictions about autonomy-supportive learning environments.

Reasoning Quality and Technical Analysis

In the subset of student–chatbot and student–student exchanges that the observers directly witnessed or that students surfaced during lecturer interaction, a recurrent pattern appeared: questions framed as 'what is the correct process?' earlier in the five weeks were, in later sessions, more often accompanied by 'why is this appropriate?' reasoning within the group. Because chatbot dialogue was observed opportunistically rather than systematically, and no message-level logs were analysed, the timing and prevalence of this shift cannot be quantified. The pattern is offered as a candidate hypothesis for the planned log-based analysis. In observed group discussions, technical analysis appeared more explicit and structured later in the five-week period than earlier, and students more frequently referenced the characteristics of a high-quality professional product when justifying process choices. Whether this reflects the chatbot, the lecturer's facilitation, peer discussion, or the combination remains an open question for subsequent analysis. The observers noted instances of students challenging their own assumptions and chatbot suggestions, asking follow-up questions, requesting clarification of AI reasoning, and cross-checking against course materials. The claim that such behaviours increased over time is based on the observers' collective impression across the five weeks, not on coded frequency data, and should be treated as hypothesis-generating. This critical stance is particularly valuable given concerns about over-reliance on AI systems.

Productive Engagement with AI Fallibility

Most significant was student engagement with AI errors. The observers repeatedly noticed instances of chatbot output that the student group themselves, or the lecturer on circulation, identified as inappropriate for the product under analysis. In the observed cases, these errors were frequently traceable to incomplete or imprecise student input – missing material specification, omitted tolerance or quantity information, or ambiguous geometric description – rather than to failures of the retrieval base. No systematic error audit was conducted, and no rate of incorrect output is claimed. When students provided vague descriptions or omitted technical details, suggestions were more likely to be inappropriate. In the observed cases, this input–output relationship appeared to function as a learning moment: students discussed what information the chatbot had been given and what had been omitted, and revised their input accordingly. Whether this generalises across groups, or to professional practice, is an open question. In the observed cases where errors were identified, students were seen to refine problem framing, improve prompt precision, verify reasoning against course materials, and cross-check with peers. The 'productive failure' and 'desirable difficulties' literatures offer a plausible theoretical reading of this pattern, but the present data do not establish that students 'developed' critical evaluation skills in a measured sense.

Understanding of Professional Quality Standards

In group discussions observed later in the five-week period, students more frequently articulated what a high-quality professional product (beroepsproduct) should contain. Without pre/post measurement or a comparison cohort, the claim that understanding itself became clearer cannot be substantiated here. The chatbot's consistent requests for justification and evidence plausibly contributed to making these quality expectations salient during group work, alongside lecturer facilitation and peer discussion. Disentangling the contribution of each is beyond what observational data support. Students were repeatedly observed articulating that professional engineering work requires explicit reasoning, not only correct answers. Whether this represents internalisation or situational performance cannot be determined from the present data. The worked examples in the knowledge base provided concrete models of professional quality, and the chatbot's questioning helped students understand what made those examples successful. This aligns with Explicit Instruction principles regarding the importance of models and worked examples for novice learners.

DISCUSSION

The patterns the observers noticed are broadly consistent with the directions the guiding theories would predict. This consistency should not be read as confirmation of any single theory, nor as evidence that the theoretical framework caused the observed patterns; observational alignment with theory is a weak test, particularly when six complementary theories are in play and the observers were not blind to the design intent.

Theoretical Interpretation of Observations

From a CLT perspective, the structured questioning approach was designed to manage cognitive load, and the observers did not notice signs of overwhelm during group work. Cognitive load was not directly measured. The Explicit Instruction elements functioned as intended: students received scaffolding and models without replacement of their own reasoning. The Socratic approach ensured students constructed understanding rather than passively receiving it. Formative Assessment principles were operationalised through continuous AI-mediated questioning that created feedback loops throughout the learning process, enabling real-time adjustment. The input–output relationship between student analysis quality and AI response quality was designed to surface metacognitive control, and the observers noted instances of students monitoring and adjusting their prompts in response to chatbot output. The claim that SRL as a capacity developed is not one the present data can support; it is a candidate hypothesis for longitudinal follow-up. SDT needs appeared wellsupported: autonomy through student control of interaction pace, competence through mastery-focused feedback, and relatedness through the collaborative learning structure. Sociocultural Theory helps explain how the chatbot functioned as a stabilising conversational partner, stimulating discussion and providing a shared reference point for negotiation of meaning.

The Instructional Ecology: Chatbot, Peers, and Lecturer

A candid interpretation of the observed shifts requires disentangling the chatbot's contribution from that of the broader instructional ecology in which it was embedded. The shift from answer-seeking to reasoning-seeking, and the growing critical evaluation of AI output over the five-week period, cannot be attributed to the chatbot alone. Three agents were at work simultaneously, and the design depended on their interaction.

The lecturer's role remained active throughout. The two lecturers circulated between groups, questioned students on their reasoning, prompted groups to share and justify their current chatbot exchange, and reminded students of their own responsibility to arrive at a defensible design decision rather than outsourcing the judgement to the system. Crucially, the chatbot's structured Socratic design also allowed the lecturers to hand off tasks that would otherwise have required direct instruction, particularly the patient, repeated questioning that surfaces missing product analysis, which no single lecturer can sustain across approximately 150 students in parallel. The instructional ecology therefore redistributed lecturer effort rather than reducing it: more time on deep reasoning with individual groups, less time on first-pass diagnostic questioning.

The peer group contributed through three concrete behaviours observed repeatedly across the five weeks: negotiation of prompts before they were submitted to the chatbot, joint evaluation of chatbot output before any of it was acted upon, and mutual explanation when one group member understood a process decision that another did not. These peer behaviours are themselves predicted by sociocultural learning theory (Vygotsky, 1978; Lave & Wenger, 1991), and the collaborative chatbot design was specifically intended to elicit them, but the peers, not the chatbot, are the ones doing this work.

The chatbot's specific contribution was to sustain a particular kind of questioning pressure that is difficult to maintain through lecturer or peer interaction alone: consistently asking for justification, surfacing missing information, and refusing to provide a finished answer. Students were not told explicitly that the chatbot was designed as a Socratic tutor; they were told only that it was not an answer machine, but a system intended to stimulate their learning. The shift toward reasoning-seeking behaviour therefore emerged from students' direct experience of the chatbot's behaviour in interaction with lecturer facilitation and peer discussion, rather than from instructional framing of the tool's purpose.

This ecological reading has two consequences for how the findings should be read. First, the observed effects should be understood as properties of a designed learning environment in which chatbot, peers, and lecturer each play a distinct role, rather than as properties of the chatbot itself. Transferring the intervention to another context therefore requires transferring all three elements, not only the technical artefact. Second, the lecturer's role in this design is not diminished by the introduction of AI – it is repositioned. Pedagogical expertise becomes more visible, not less, precisely because the chatbot takes over the repetitive scaffolding work and frees the lecturer to intervene where disciplinary judgement is most needed.

Central Insight: AI as Feedback Instrument

AI is educationally valuable when designed as a structured feedback instrument rather than an answer engine. AI tools that provide direct answers may satisfy immediate student needs but undermine learning by removing productive struggle. AI tools designed as feedback instruments—that question, probe, and guide rather than tell—maintain cognitive engagement and support deeper learning. This has implications for the broader AI-in-education field. The theoretical framework here provides principles for distinguishing educationally valuable AI applications from potentially harmful ones.

Implications for the CDIO Community

This work has several implications for the CDIO community. First, it demonstrates that AI can be integrated into CDIO programmes in ways that strengthen rather than undermine core educational values through thoughtful design grounded in learning theory. Second, it provides a model for how AI tools can support formative feedback and active learning emphasised in CDIO Standards 8 and 11, extending feedback opportunities beyond what instructors can provide individually. Third, preparing students for professional AI use can be integrated with core disciplinary learning. Students simultaneously learn manufacturing process design and develop critical AI evaluation skills. Recent CDIO work supports these findings. Nikolic et al. (2025) developed a CDIO-based framework for AI integration, Lei et al. (2025) demonstrated positive effects of generative AI-enabled CDIO teaching on computational thinking, and Karoui and Bettaieb (2024) presented strategies for AI-supported active learning in engineering education.

CONCLUSIONS

This study has described a theory-informed Socratic chatbot intervention in a first-year manufacturing course, grounded in six complementary learning theories and enacted through an instructional ecology in which chatbot, peers, and lecturer each play a distinct role. Observational evidence across five weeks and four class groups is consistent with the theoretical predictions: students shifted from answer-seeking to reasoning-seeking, engaged productively with AI fallibility, and developed a clearer sense of what a high-quality professional artefact demands. These claims are offered cautiously, as candidate design hypotheses to be tested in subsequent DBR cycles through the survey and inter-cohort analysis now in preparation.

The contribution to CDIO Standards 2 and 7 is where the pedagogical development lands most concretely. Standard 2 is strengthened by aligning the course's learning outcomes with the AI-literacy signals now converging from industry and society — the World Economic Forum, FME, Koninklijke Metaalunie, the SER, and the EU AI Act — and mapping these onto CDIO Syllabus 3.0 Sections 2.4, 2.5, and 3.1–3.2, alongside the disciplinary core. Standard 7 is strengthened by designing an activity that mirrors a recognisable category of contemporary engineering work: the production engineer using GenAI to triage design-for-manufacturing issues, integrating process choices, design choices, and responsible AI use within a single task. Students perform a legitimate peripheral version of this workflow rather than simulating it at a distance.

A broader design principle follows: the general-purpose large language models supplied by commercial AI providers are not, by themselves, learning companions. They become learning partners only when augmented with pedagogy and disciplinary knowledge — through curriculum-grounded retrieval, behavioural instructions derived from learning theory, and embedding within a classroom ecology of peers and lecturers. Transferring this intervention therefore means transferring the ecology, not only the artefact. Future work will extend the evidence base through systematic student evaluation, inter-cohort comparison, and longitudinal follow-up, and the CDIO community is invited to replicate and challenge these design principles in other disciplinary contexts.

ACKNOWLEDGEMENTS

The author thanks Eugenio Tamassia, the second lecturer and co-observer, for his collaboration in the parallel class groups and for sharing his observations that informed this study. Sincere thanks are also extended to all students of Productietechniek 1 who participated in the chatbot-supported sessions and whose engagement made this design-based research cycle possible. Finally, the author is grateful to Estella Griffioen and Manika Garg for their thoughtful review of earlier drafts of this manuscript.

The author received no financial support for this work.

AI usage declaration

AI tools (ChatGPT, Claude and Gemini) were used both as part of the educational intervention described in this paper and for language editing and drafting support in preparing this manuscript. All conceptual, theoretical, and pedagogical design decisions were made by the author. The author takes full responsibility for the content and any errors or omissions.

REFERENCES

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, 41(1), 16–25. https://doi.org/10.3102/0013189X11428813

Archer, A. L., & Hughes, C. A. (2011). Explicit instruction: Effective and efficient teaching. Guilford Press.

Bakker, A. (2018). *Design research in education: A practical guide for early career researchers*. Routledge. https://doi.org/10.4324/9780203701010

Batista, J., Mesquita, A., et al. (2024). Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review. Information, 15(11), Article 676. https://doi.org/10.3390/info15110676

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74. https://doi.org/10.1080/0969595980050102

CDIO Initiative. (2020). CDIO Standards 3.0. Worldwide CDIO Initiative. https://cdio.org/files/CDIO%20STANDARDS%203.pdf

Crawley, E. F., Malmqvist, J., Östlund, S., Brodeur, D. R., & Edström, K. (2014). Rethinking engineering education: The CDIO approach (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-05561-9

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Danyaro, K. U., Abdullahi, S., Abdallah, A. S., & Chiroma, H. (2025). Hallucinations in large language models for education: Challenges and mitigation. *International Journal of Teaching, Learning and Education*, 4(6), 639993. https://doi.org/10.22161/ijtle.4.6.2

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. Plenum Press. https://doi.org/10.1007/978-1-4899-2271-7

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Edström, K., & Kolmos, A. (2014). PBL and CDIO: Complementary models for engineering education development. European Journal of Engineering Education, 39(5), 539–555. https://doi.org/10.1080/03043797.2014.895703

Edström, K., Malmqvist, J., & Roslöf, J. (2020). Scholarly development of engineering education—the CDIO approach. European Journal of Engineering Education, 45(1), 1–3. https://doi.org/10.1080/03043797.2020.1704361

Favero, L., Pérez-Ortiz, J.A., Käser, T., & Oliver, N. (2024). Enhancing critical thinking in education by means of a Socratic chatbot. Proceedings of the International Workshop on AI in Education and Educational Research, ECAI 2024. https://arxiv.org/abs/2409.05511

FME. (2025). *Smart Industry Productiviteitsagenda 2026–2028*. FME / Smart Industry. https://www.fme.nl

Hansen, G., & Sindre, G. (2023). Assessment and feedback across various outcomes in project-courses: A department-wide study. Proceedings of the 19th International CDIO Conference (pp. 913–922). NTNU, Trondheim, Norway.

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge. https://doi.org/10.4324/9780203887332

Hon, K. (2025). Generative AI in higher education: A systematic review of its effects on learning outcomes and academic performance. Journal of Educational Computing Research. https://doi.org/10.1177/00472395251400089

Hughes, C. A., Morris, J. R., Therrien, W. J., & Benson, S. K. (2017). Explicit instruction: Historical and contemporary contexts. Learning Disabilities Research & Practice, 32(3), 140–148. https://doi.org/10.1111/ldrp.12142

Karoui, S., & Bettaieb, L. (2024). AI-supported active learning in engineering education: A comprehensive teaching strategy. Proceedings of the 20th International CDIO Conference. ESPRIT, Tunis, Tunisia.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. Review of Educational Research, 86(1), 42–78. https://doi.org/10.3102/0034654315581420

Labadze, L., Grigolia, M. & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. International Journal of Educational Technology in Higher Education, 20, 56. https://doi.org/10.1186/s41239-023-00426-1

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge University Press.

Lei, X., Liu, J., Fu, X., Zhao, J., & Yi, B. (2025). The effects of a generative AI-enabled CDIO teaching model on undergraduates' computational thinking and individual psychological constructs. Computer Applications in Engineering Education. https://doi.org/10.1002/cae.70075

Liang, J., Stephens, J. M., & Brown, G. T. L. (2025). A systematic review of the early impact of artificial intelligence on higher education curriculum, instruction, and assessment. Frontiers in Education, 10, Article 1522841. https://doi.org/10.3389/feduc.2025.1522841

Liu, Z., Agrawal, P., Singhal, S., Madaan, V., Kumar, M., & Verma, P. K. (2025). LPITutor: An LLM based personalized intelligent tutoring system using RAG and prompt engineering. PeerJ Computer Science, 11, e2991. https://doi.org/10.7717/peerj-cs.2991

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. Journal of Educational Psychology, 106(4), 901–918. https://doi.org/10.1037/a0037123

Malmqvist, J., Edström, K., & Rosén, A. (2020). CDIO Standards 3.0—Updates to the core CDIO standards. Proceedings of the 16th International CDIO Conference. Chalmers University of Technology, Gothenburg, Sweden.

Malmqvist, J., Knutson Wedel, M., Lundqvist, U., Edström, K., Rosén, A., Fruergaard Astrup, T., Vigild, M., Munkebo Hussmann, P., Grøm, A., Lyng, R., Gunnarsson, S., Leong, H., & Kamp, A. (2019). Towards CDIO Standards 3.0. Proceedings of the 15th International CDIO Conference. Aarhus University, Aarhus, Denmark.

Malmqvist, J., Lundqvist, U., Rosén, A., Edström, K., Gupta, R., Leong, H., Cheah, S. M., Bennedsen, J., Hugo, R., Kamp, A., Leifler, O., Gunnarsson, S., Roslöf, J., & Spooner, D. (2022). The CDIO Syllabus 3.0: An updated statement of goals. Proceedings of the 18th International CDIO Conference (pp. 18–36). Reykjavík University.

McKenney, S., & Reeves, T. C. (2018). *Conducting educational design research* (2nd ed.). Routledge. https://doi.org/10.4324/9781315105642

Memarian, B., & Doleck, T. (2024). Generative AI chatbots in higher education: A review of an emerging research area. Higher Education. https://doi.org/10.1007/s10734-024-01288-w

Nikolic, S., Quince, Z., Lidfors Lindqvist, A., Neal, P., Grundy, S., Lim, M., Tahmasebinia, F., Rios, S., Burridge, J., Petkoff, K., Chowdhury, A. A., Lee, W. S. L., Prestigiacomo, R., Fernando, H., Lok, P., & Symes, M. (2025). Project-work Artificial Intelligence Integration Framework (PAIIF): Developing a CDIO-based framework for educational integration. STEM Education, 5(2), 310–332. https://doi.org/10.3934/steme.2025016

Ofosu-Ampong, K. (2024). A systematic review of generative AI for teaching and learning practice. Education Sciences, 14(6), Article 636. https://doi.org/10.3390/educsci14060636

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. Frontiers in Psychology, 8, Article 422. https://doi.org/10.3389/fpsyg.2017.00422

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology, 82(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Rathenau Instituut. (2023). *Rathenau Scan Generatieve AI*. Rathenau Instituut.

Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. American Educator, 36(1), 12–19, 39.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 376–391). Macmillan.

Ruthven, K. (2013). The calculator debate. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 431–457). Springer.

Ryan, R. M. (Ed.). (2023). The Oxford handbook of self-determination theory. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197600047.001.0001

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68–78. https://doi.org/10.1037//0003-066x.55.1.68

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. Instructional Science, 18(2), 119–144. https://doi.org/10.1007/BF00117714

Sociaal-Economische Raad. (2025). *AI en werk: Samen naar een werkende toekomst met AI* (Advies 25/03). SER. https://www.ser.nl/nl/adviezen/ai-en-werk

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction, 4(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251–296. https://doi.org/10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. Educational Psychology Review, 31(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. Humanities and Social Sciences Communications, 12, Article 621. https://doi.org/10.1057/s41599-025-04787-y

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

World Economic Forum. (2025). *The Future of Jobs Report 2025*. World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2025/

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. British Journal of Educational Technology, 55(3), 1051–1066. https://doi.org/10.1111/bjet.13334

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

Zhang, L., et al. (2024). SPL: A Socratic playground for learning powered by large language model.. DOI:10.48550/arXiv.2406.13919

Zhang, Z., et al. (2024). A systematic review on prompt engineering in large language models for K-12 STEM education. arXiv preprint arXiv:2410.11123. https://arxiv.org/abs/2410.11123

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. Journal of Educational Psychology, 81(3), 329–339. https://doi.org/10.1037/0022-0663.81.3.329

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of self-regulation (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

BIOGRAPHICAL INFORMATION

Abboy Verkuilen is a senior Lecturer in Mechanical Engineering and Researcher Smart Sustainable Manufacturing, focusing on leveraging (Generative) AI both at The Hague University of Applied Sciences, specializing in engineering education and the integration of technology in learning. As co-author of *Basisboek Productietechniek* (with Tempelman & Eissen), he has contributed significantly to manufacturing education. He is also co-author of the 4th print of the book Ontwerken van Technische Innovaties, where focused on leveraging GenAI in the innovation process. He holds an Industrial Design Engineering degree from Delft University of Technology. When not teaching future engineers, you'll find him exploring new ways to blend technology with effective pedagogy. Abboy designs innovative educational experiences that blend theory with practical application, fostering student engagement and mastery. His current work includes an e-learning platform empowering engineering students to use AI effectively in their studies. He was nominated for Teacher of the Year 2020-2021 at The Hague University of Applied Sciences. Abboy has presented workshops on AI-powered personalized learning at international conferences including Online Educa Berlin, SEFI and EFYE.

Corresponding author

Abboy Verkuilen MSc
THUAS The Hague University of Applied Sciences
Faculty of Technology, Innovation & Society
Department of Mechanical Engineering
Rotterdamseweg 137
2628AL Delft, The Netherlands
A.Verkuilen@hhs.nl