



AI Tools for Instructional Design: Analyzing the Gap Between Technological Potential and Pedagogical Rigor

Abdelmounaim Azinda¹, Mohamed Khaldi²

Laboratory of Information Technologies and System Modeling, Abdelmalek Essaadi University, Faculty of Science, Tetouan, Morocco¹

Laboratory of Information Technologies and System Modeling, Abdelmalek Essaadi University, Faculty of Science, Tetouan, Morocco²

abdilmounaimazinda2000@gmail.com

ABSTRACT

Dedicated generative AI applications are increasingly targeting the Instructional Design (ID) sector, claiming to automate tasks ranging from lesson planning to assessment creation. This descriptive study distinguishes these specialized "vertical" tools from general-purpose Large Language Models to evaluate their actual utility. We contrasted the functional output of representative platforms (specifically MagicSchool.ai, CourseAI, and IDEAI) against established ID foundations: needs analysis, Constructive Alignment, and Cognitive Load Theory. The analysis identifies two primary tool categories: Global Design Assistants and Content Generators. While these systems demonstrate high efficiency in drafting initial structures, results indicate substantial limitations regarding pedagogical rigor. The tools generally fail to conduct strategic needs assessments and struggle to generate assessments that reach the higher levels of Bloom's Taxonomy. Furthermore, the alignment between objectives and activities often appears mechanical rather than conceptual. We conclude that while such AI tools are viable for accelerating specific production tasks, they cannot currently replace the strategic decision-making of human experts. Effective implementation therefore requires a human-in-the-loop approach to validate the relevance and depth of the generated materials.

Keywords: Instructional Design; Generative AI; Educational Technology; Constructive Alignment; Human-AI Collaboration; Pedagogical Engineering.

1. Introduction

1.1 Context: General vs. Specialized AI in Education

Generative Artificial Intelligence (GenAI) models, such as ChatGPT and similar conversational agents, have been widely adopted in the educational sector. Initial adoption focused on general-purpose tasks: drafting text, brainstorming ideas, or summarizing documents (Kasneji et al., 2023). Professional practice is now shifting. Beyond these general tools, a distinct market of specialized applications is developing specifically for Instructional Design (ID). Unlike standard chatbots, these new platforms (e.g., MagicSchool, CourseAI, IDEAI) are integrated directly into authoring workflows or Learning Management Systems (LMS). They do not simply generate text; they claim to structure entire learning experiences.

1.2 The Technical Specificity of the New Tools

This study distinguishes "vertical" AI tools from general models. These specialized agents offer interfaces designed for the ID workflow, proposing to automate steps such as objective formulation, content mapping, and assessment creation. This evolution represents a move from adaptive use of generic tech to the adoption of dedicated ID

software. The promise is operational efficiency: reducing the time required to produce course drafts (Mollick & Mollick, 2022).

1.3 Problem Statement: Efficiency Versus Pedagogical Standards

The availability of these tools raises a validation issue. Increased production speed does not guarantee instructional quality. Instructional Design is a systematic process dependent on specific theoretical frameworks, not just content production (Reigeluth, 1999). A core problem remains: the gap between the capability to generate educational artifacts and the scientific rigor required for effective learning. It is unclear if these tools adhere to fundamental principles like Needs Analysis (Rossett, 1987) or Constructive Alignment (Biggs & Tang, 2011). If tools prioritize ease of generation over these methodological requirements, there is a risk of producing superficially coherent but pedagogically ineffective materials (Kirschner, Sweller, & Clark, 2006).

1.4 Study Objectives

Existing literature largely discusses the potential of GenAI in broad terms. There is a lack of critical analysis specifically auditing these vertical ID tools against standard design principles. This paper addresses this gap. The primary objective is to evaluate the reliability of specialized AI assistants in supporting a rigorous ID process. We examine their typology, functionality, and intrinsic limitations by comparing their outputs with established instructional engineering criteria, specifically referring to Bloom's Taxonomy and principles of cognitive load.

2. Literature Review and Theoretical Framework

2.1 Generative AI in the Educational Context

Recent studies indicate that AI tools offer utility for task automation and personalized feedback (Kasneji et al., 2023). In particular, Generative AIs (GenAI) facilitate rapid content production (Mollick & Mollick, 2022). Despite these capabilities, substantial limitations are documented in the literature. Issues regarding the accuracy of generated information, the presence of algorithmic bias, and data privacy remain significant concerns (Selwyn, 2022). Furthermore, there is a noted tendency for these models to produce generic outputs lacking critical depth (Gao et al., 2023).

2.2 Instructional Design (ID) Foundations

This study evaluates AI tools against established ID standards. Quality instructional design is defined as a systematic process to optimize learning (Reigeluth, 1999). It relies on specific theoretical requirements:

- **Needs Analysis:** Investigating the learner characteristics and performance gaps is the initial step for relevant design (Rossett, 1987). Design decisions must be based on context, not just content availability.
- **Taxonomy of Objectives:** Effective instruction requires clear objectives. Bloom's Taxonomy and its revisions provide a hierarchical structure, distinguishing between lower cognitive levels (Knowledge) and higher levels such as Analysis and Evaluation (Anderson & Krathwohl, 2001; Bloom et al., 1956).
- **Constructive Alignment:** This principle dictates that learning activities and assessment tasks must align directly with the stated learning objectives (Biggs & Tang, 2011; Wiggins & McTighe, 2005). Coherence between these elements is necessary for valid assessment.
- **Instructional Strategy and Cognitive Load:** Design choices should facilitate learner engagement (Merrill, 2002). Furthermore, managing the cognitive load is required to prevent overwhelming the learner's working memory during instruction (Sweller, van Merriënboer, & Paas, 2019).
- **Iterative Process:** Frameworks such as ADDIE imply a cycle of evaluation and revision (Branch, 2009). The process is rarely linear.

2.3 The Gap: AI Capabilities vs. ID Requirements

Existing research has primarily examined general AI applications. There is limited documentation specifically analyzing the compliance of specialized "ID chatbots" with the theoretical principles listed above. The central question of this study concerns the operational capability of these tools. It is necessary to determine if current tools can perform complex tasks—such as Constructive Alignment and Contextual Analysis—or if they are limited to surface-level content generation.

3. Methodology

3.1 Study Design

This study applies a descriptive qualitative analysis. The methodological approach consists of a functional audit of software capabilities relative to Instructional Design (ID) principles. The goal is to evaluate technical features against theoretical requirements rather than to measure user satisfaction levels.

3.2 Sample Selection and Data Sources

Data collection was performed between January and March 2024. The selection of tools followed a purposeful sampling strategy to identify "Specialized ID AI" distinct from general Large Language Models.

The inclusion criteria were:

1. **Market positioning:** The tool explicitly claims to assist instructional designers or educators.
2. **Functionality:** The platform offers features for course structuring, assessment creation, or lesson planning.
3. **Availability:** The tool was publicly accessible for testing during the study period.

Selected platforms for the analysis include *MagicSchool*, *CourseAI*, *IDEAI*, *QuestionWell*, and *Khanmigo*. The full inventory is detailed in Appendix A. Data sources included technical documentation, vendor white papers, and direct functional testing of the interfaces.

3.3 Analytical Procedure

The evaluation was conducted in four distinct stages:

- **Phase 1: Typological Categorization.** Tools were classified according to their primary function (e.g., Content Generator vs. Process Assistant) and intended user base (K-12 teachers vs. ID professionals).
- **Phase 2: Controlled Testing (Prompting).** A "Black Box" testing protocol was applied. Standardized prompts were entered into each platform to generate specific instructional artifacts. Test scenarios included:
 - *Scenario A (Simple):* Generate a lesson outline on a general topic.
 - *Scenario B (Complex):* Design a problem-solving activity for a specific learner profile with constructive alignment constraints.
- **Phase 3: Theoretical Evaluation.** Generated outputs were audited against the framework defined in Section 2. Coding criteria focused on:
 - Presence of learner analysis mechanisms.
 - Cognitive level of generated questions (based on Bloom's Taxonomy).
 - Alignment accuracy between stated objectives and proposed assessments.
- **Phase 4: Synthesis.** Observed limitations were documented and cross-referenced with ID literature standards to determine reliability.

3.4 Methodological Scope

This study evaluates the baseline capabilities of the software ("Zero-shot" performance). It does not account for advanced output manipulation by expert users. Results are descriptive of the technology state at the time of analysis and are subject to change with software updates.

4. Results

4.1 Functional Classification of Tools

Data analysis establishes the existence of a distinct software category, differing from general generative AI. These platforms incorporate "Interface Scaffolding": the user interacts via preset form fields rather than open prompting. Two primary typologies were identified within the sample:

- **Structure and Sequence Generators:** Tools like *CourseAI* or *MagicSchool* focus on producing global artifacts (syllabi, lesson plans, course outlines).
- **Granular Asset Generators:** Tools like *QuestionWell* or *Diffit* focus on specific micro-tasks (generating assessment items, levelling reading text).

4.2 Deficiency in Needs Analysis (The "Context Gap")

A systematic functional limitation was observed regarding the "Analysis" phase of the ADDIE framework. In 100% of tested scenarios, the tools initiated content generation immediately after the prompt was received. No platform requested clarification on learner demographics, prior knowledge, or performance gaps before generating output. The interaction model is strictly "Output-Oriented." The software assumes the validity of the user's request without diagnosing the pedagogical need, effectively skipping the strategic context analysis required for rigorous design.

4.3 Limitations in Cognitive Complexity (Bloom's Taxonomy)

The content generated by the tools exhibited a bias toward lower-order cognitive skills. Analysis of generated assessment items revealed a predominance of "Remember" and "Understand" tasks.

- **Observation:** When prompted to create questions for "Conflict Resolution" (a High-Level, soft-skill topic), the tools produced multiple-choice questions focusing on vocabulary definitions.
- **Interpretation:** The tools struggled to autonomously design activities requiring "Evaluation" or "Creation." Generating authentic performance tasks (simulations, complex case studies) remains unreliable without extensive manual modification by a human expert.

4.4 The Mechanical Nature of Alignment

Tests evaluated the connection between Learning Objectives and generated Assessments. While the tools demonstrated "Semantic Alignment" (using similar keywords in both the objective and the quiz), they often failed to achieve "Constructive Alignment" (matching the cognitive demand).

- **Failure Mode:** An objective specifying "Analyze the root causes" was frequently paired with a true/false question.
- **Result:** The generated artifacts present a structural cohesion that masks a lack of pedagogical validity. The alignment is syntactic, not functional.

4.5 Production Efficiency vs. Instructional Value

The tools performed reliably regarding speed and format structure. Outputs were grammatically correct and logically formatted (headers, bullet points). However, the pedagogical "density" was variable. To satisfy prompt requirements

for length, tools frequently inserted generic activities (e.g., word searches) unrelated to the learning goals, increasing extraneous cognitive load without contributing to the learning outcome.

5. Discussion

5.1 The Industrialization of Instructional Output

The data indicate a fundamental tension in the current tool landscape. While the audited platforms successfully reduce the temporal cost of artifact production, they promote a linear "Input-Output" workflow that contradicts the recursive nature of Instructional Design (Branch, 2009). By bypassing the analysis phase, these tools institutionalize a "Solution-First" bias. The widespread adoption of such "proxy-designers" risks shifting the industry focus from *problem-solving* (identifying the correct performance intervention) to *content manufacturing* (producing courses regardless of need). This phenomenon constitutes a regression toward behaviorist content transmission, prioritizing volume over diagnostic accuracy.

5.2 The Risk of Professional De-Skilling

The dominance of "scaffolded interfaces" (black-box templates) raises concerns regarding practitioner competency. Novice designers relying on automated alignment suggestions may fail to develop the necessary schemas for "Constructive Alignment." If the alignment logic is outsourced to an algorithm that privileges semantic similarity over cognitive congruency (as observed in the *Conflict Resolution* test case), the designer operates without understanding the underlying pedagogical mechanics. This creates a dependency loop, potentially eroding the critical analytical skills required to diagnose instructional failures manually.

5.3 Epistemic Validity and the "Contextual Blindness" Problem

The limitation regarding Bloom's Taxonomy signals an intrinsic ceiling to current automation. The tools function as statistical engines, forcing convergence toward the mean—specifically, the lower cognitive levels of definition and recall. They lack the ontological understanding to model ambiguity or high-stakes contexts (e.g., medical ethics, safety protocols) without explicit human guidance. Therefore, the "autonomous instructional designer" is currently a marketing construct, not a technological reality.

5.4 Paradigm Shift: From Creation to Auditing

Given the identified inability of vertical AI to execute strategic alignment independently, the role of the instructional designer must evolve. The value proposition shifts from the *generation* of materials to the *validation* of materials. This requires a transition to an "Audit-Based" workflow. In this model, the AI serves as a high-volume drafter, while the human expert functions as the "Epistemic Gatekeeper," specifically responsible for:

1. Injecting contextual constraints that the AI ignores.
 2. Elevating the cognitive complexity of generated assessments.
 3. Verifying the functional isomorphism between objectives and activities.
- Dependence on AI without this verification layer violates the core standards of evidence-based pedagogy.

6. Conclusion

6.1 Verdict on Algorithmic Reliability

This heuristic audit provides a technical rebuttal to the marketing claims surrounding autonomous Instructional Design agents. The evidence establishes a clear functional separation: while vertical AI tools demonstrate high competence in Syntactic Structuring (formatting, outlining, asset generation), they exhibit a systematic failure in Pedagogical Reasoning (alignment, diagnosis, complexity management). The tools operate as effective "accelerants"

for administrative tasks but lack the teleological capacity to ensure learning transfer. Speed of production has been achieved at the expense of strategic analysis.

6.2 The Mandatory Pivot to "Quality Assurance"

Consequently, the model of "unsupervised automation" is rejected for professional practice. The "Context Gap" identified in the Results necessitates a structural change in the Instructional Designer's mandate. The professional role transforms from a *Builder* of content to an *Auditor* of algorithmic logic. The human expert is the only component of the system capable of validating the "Constructive Alignment," ensuring that the efficiency of the machine does not result in the vacuity of the instruction.

6.3 Implications for Future Research

The burden of proof now shifts from *Software Functionality* to *Learning Efficacy*. Future inquiry must move beyond descriptive analysis of tool features to rigorous empirical testing of student outcomes. Unless AI-generated courses can demonstrate learning retention rates equivalent to human-designed curricula, the efficiency gains remain technically impressive but educationally irrelevant. Efficiency is a manufacturing metric, not a learning metric.

References

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). McGraw-Hill Education.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay Company.
- Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer.
- Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (4th ed.). Wiley.
- Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, 6(4), 50–72. <https://doi.org/10.1111/j.1937-8327.1993.tb00605.x>
- Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., et al. (2023). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *BioRxiv*. <https://doi.org/10.1101/2023.01.26.525805>
- Kasneji, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
- Mager, R. F. (1997). *Preparing instructional objectives: A critical tool in the development of effective instruction* (3rd ed.). The Center for Effective Performance.
- Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, 50(3), 43–59. <https://doi.org/10.1007/BF02505024>
- Mollick, E. R., & Mollick, L. (2022). New modes of learning enabled by AI chatbots: Three methods and assignments. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4300783>
- Perkins, D. N., & Salomon, G. (1992). Transfer of learning. *International Encyclopedia of Education*, 2, 6452–6457.
- Reigeluth, C. M. (Ed.). (1999). *Instructional-design theories and models: A new paradigm of instructional theory* (Vol. 2). Lawrence Erlbaum Associates.
- Rossett, A. (1987). *Training needs assessment*. Educational Technology Publications.
- Selwyn, N. (2022). EdTech algorithms: Deploying intelligent technologies in education. *Discourse: Studies in the Cultural Politics of Education*, 43(5), 814–827. <https://doi.org/10.1080/01596306.2022.2105868>

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>

Wiggins, G. P. (1990). The case for authentic assessment. *Practical Assessment, Research & Evaluation*, 2(2). <https://doi.org/10.7275/fb1-mm19>

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). Association for Supervision and Curriculum Development (ASCD).

Appendix A

Inventory of Specialized Software Analyzed (Jan-Mar 2024)

A.1 Sampling Criteria

The tools listed below were selected based on their availability in the standard EdTech market as of Q1 2024. We prioritized platforms specifically marketed for instructional design tasks over general-purpose Large Language Models.

A.2 Conflict of Interest Statement

The authors confirm no commercial relationship with any of the entities listed below. This list is provided for replication purposes only.

Table A1

Operational Characteristics of the Study Corpus

Tool Name	Functional Category	Usage in Study (Scope)	Official Access Point
MagicSchool.ai	Teaching Assistant Suite	Utilized to test automated lesson planning and rubric generation for K-12 contexts.	magicschool.ai
CourseAI	Course Prototyping	Used to evaluate rapid curriculum structuring from minimal keyword inputs.	courseai.com
Khanmigo	Socratic Tutor (Pilot)	Tested for real-time feedback loops and student-facing interactions.	khanacademy.org/khan-labs
IDEAI (dominKnow)	Integrated Authoring Aid	Examined for its capacity to suggest content improvements within a specialized LMS authoring tool.	dominKnow.com
QuestionWell	Assessment Generator	Analyzed specifically for alignment between input text and generated multiple-choice questions.	questionwell.org
Curipod	Interactive Lesson Tool	Evaluated for its ability to create immediate engagement slides based on learning objectives.	curipod.com
Diffit	Differentiation Engine	Tested on its ability to re-level complex text for different reading capabilities.	web.diffit.me

Tome	Presentation Generator	Used to assess the coherence between visual slides and instructional text narratives.	tome.app
Cognii	Formative Assessment	Included to verify capabilities in open-ended answer analysis using NLP.	cognii.com
Articulate 360 AI	e-Learning Assistant	Included as a benchmark for AI features embedded in industry-standard software.	articulate.com

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

