Multimodal Cognitive Digital Twins: Modeling the Impact of Misleading Charts and Texts in School-Based Learning

Vasyl Zalizko*

ETH AI Centr, Zurich, Switzerland

ARTICLE INFO

Keywords:
Multimodal learning analytics
Cognitive digital twins
Data visualization education
Human Multimodal Misleadingness Index
Misinformation resilience
Graph literacy
OCR-based handwritten text recognition.

ABSTRACT

Misleading information in educational settings rarely takes a single form. Learners may be misled by distorted graphs, biased framing, omissions, suggestive wording, and situations in which visual and textual cues reinforce one another. Unlike earlier studies that focused on graphical or textual materials alone, this study examines misleadingness at three levels: visual, textual, and mixed chart–text. To capture this broader phenomenon, we introduce a unified framework of multimodal cognitive digital twins and a composite measure, the Human Multimodal Misleadingness Index (HMMI). Unlike binary scoring, this index provides a more fine-grained description of learners' understanding of potentially misleading materials and their resistance to being misled. In the empirical implementation used here, higher HMMI values indicate stronger learner understanding and greater resistance to misleading cues. The study was conducted as a fully online school-based intervention with learners from an international gymnasium. The three instructional groups were broadly comparable at pretest. After the intervention, the strongest improvement was observed in the full digital-twin condition, with a mean HMMI gain of 0.347, followed by the standard-feedback condition (0.199), whereas the partial digital-twin condition showed the weakest improvement (0.077). Younger learners also showed lower baseline values on the visual subindex than older learners (0.561 vs. 0.687), indicating lower resilience to misleading visual structures, including visual illusions. Methodologically, the study shows that OCR- and handwriting-based analysis of learner responses enables a richer interpretation of authentic written traces, including omissions, partial understanding, and revision behavior. Overall, the findings suggest that multimodal misleadingness can be modeled more effectively when visual, textual, and mixed misleadingness are studied jointly and when learner support moves beyond binary correctness toward adaptive, cognitively informed feedback.

1. Introduction

Digital information environments rarely mislead learners through a single modality alone. In authentic educational and media contexts, misleadingness is typically produced by the interaction of textual framing, visual scaling, selective contextualization, omitted background information, and attention-guiding interface cues. Recent benchmark and empirical studies have examined how multimodal models respond to misleading charts (Rho et al., 2026; Bharti et al., 2024; Driessen et al., 2022) and how language models reason about the mental states of story characters in extended textual contexts (Zhou et al., 2025; Dai and Brell-Cokcan, 2022). However, these research traditions remain largely disconnected, even though learners are commonly exposed to multimodal materials in which visual and verbal cues jointly shape misunderstanding.

This separation is problematic because human misunderstanding is strongly context-dependent even at the perceptual level. Classic geometric and size illusions demonstrate that visual interpretation is not a direct readout of physical stimulus properties. For example, in the Ebbinghaus illusion, the same central circle appears smaller when surrounded by larger circles and larger when surrounded by smaller circles (Todorović and Jovanović, 2018; Mruczek et al., 2015; Gold, 2014). More broadly, geometric-optical illusions such as Ebbinghaus, Müller–Lyer, Ponzo, Delboeuf, and Poggendorff show that perceived size, length, position, and proportion can be systematically distorted by contextual structure (Shapiro and Todorovic, 2017). From an educational perspective, these findings are not merely of theoretical interest. They suggest that visual interpretation is inherently vulnerable to contextual manipulation and that learner-facing charts, diagrams, and infographics can induce predictable biases even when the underlying data remain unchanged.

A parallel vulnerability exists in language. Textual misleadingness does not depend only on outright falsehood. It may arise through framing, selective omission, suggestive wording, causal insinuation, and perspective control. Classic work on the framing of decisions showed that equivalent information can lead to systematically different judgments depending on how options are described (Tversky and Kahneman, 1981). Likewise, research on language and memory demonstrated that suggestive wording can alter recollection and confidence, indicating that verbal formulation can actively reshape interpretation rather than merely describe it (Loftus and Palmer, 1974). In contemporary digital environments, this matters because learners often encounter chart-like quantitative claims embedded in headlines, captions, annotations, and short persuasive narratives. Thus, visual and textual manipulations often operate jointly.

Importantly, the contemporary literature on misleading visualizations increasingly supports this broader view.

*Corresponding author.
zvasyl@eth.ch (V. Zalizko)
0798697120 (V. Zalizko)
ORCID(s): https://orcid.org/0000-0001-5362-8270 (V. Zalizko)

Recent educational work distinguishes between misleading graphical elements and misleading contextual elements, showing that titles, annotations, surrounding text, and framing cues can amplify or even stabilize false interpretations of visualizations (Rho and Rau, 2025; Rho et al., 2026). At the same time, empirical comparisons across visualization types suggest that some graphical manipulations, especially axis-related distortions, remain deceptive even for learners with relatively strong data literacy (Rho et al., 2026). This means that educationally relevant misleadingness is not reducible to a single faulty graph feature; rather, it often emerges from multimodal interaction between perceptual bias, prior expectations, and contextual framing.

This observation has direct implications for the design of digital learner models. If learners can be misled by visual context in ways analogous to classic perceptual illusions, and if their judgments can also be steered by framing and suggestive language, then a useful educational model must represent more than correctness alone. It must approximate the pathways by which learners arrive at confident but false interpretations. In other words, what is needed is not only a benchmark of model performance, but a cognitively informed representation of learner vulnerability across modalities.

In contrast to other research (Ratchita and Anandhi, 2026; Bachmann et al., 2024; Dai and Brell-Cokcan, 2022), the present study proposes a multimodal cognitive digital twin explicitly centered on student activity and learning behavior. In our framework, the digital twin does not merely replicate performance outcomes; rather, it models systematic patterns of human misunderstanding across multiple modalities, especially visual, textual, and combined chart–text representations. A central methodological component of this framework is OCR- and handwriting-based learner-response analysis, which makes it possible to incorporate authentic written learner traces into the model rather than relying only on final-answer correctness (Zalizko, 2025). In this sense, the digital twin is conceived as a multidimensional model that connects virtual and physical learning spaces through continuous evidence from student responses, behavioral traces, and contextual learning data.

More specifically, we define the multimodal cognitive digital twin as a predictive and diagnostic instrument that combines detailed analysis of students' written work with indicators of cognitive and psycho-emotional state. Its purpose is not only to estimate achievement, but also to identify groups of students who are likely to be misled, to determine the mechanism through which misleadingness occurs, to estimate the degree of confidence with which an incorrect interpretation is held, and to specify the type of support under which such misconceptions can be corrected. In this way, the proposed framework extends conventional learner modeling toward an interpretable, intervention-oriented architecture for school-based adaptive support. To operationalize this framework, we introduce the Human Multimodal Misleadingness Index (HMMI). Unlike in some related studies, the scalar HMMI value in the present study is defined on the unit interval $[0, 1]$, with higher values indicating fewer learner errors and therefore stronger understanding of misleading materials This makes it possible to move beyond binary accuracy and toward a richer educational account of misleadingness: not simply whether a learner was wrong, but how strongly, how confidently, and how durably the learner was drawn toward an incorrect interpretation.

2. Theoretical background and literature review

2.1. Perceptual foundations of misleadingness

Human interpretation of information is inherently context-sensitive. Classic work on visual illusions shows that perception is not a direct reflection of physical stimulus properties, but the result of context-dependent cognitive processing. In the Ebbinghaus illusion, for example, the same central circle appears smaller when surrounded by larger circles and larger when surrounded by smaller circles. More broadly, geometric-optical illusions such as Ebbinghaus, Müller–Lyer, Ponzo, Delboeuf, and Poggendorff demonstrate that perceived size, length, proportion, and position can be systematically distorted by surrounding visual structure (Shapiro and Todorovic, 2017; Todorović and Jovanović, 2018; Mruczek et al., 2015; Gold, 2014).

A retrospective analysis Table 1 revealed that, from an educational standpoint, these findings are significant because they highlight a fundamental principle: interpretation is influenced by both the object itself and the relational context in which it is presented. This principle extends directly to charts, diagrams, and other educational visualizations.

2.2. Misleading charts and educational visualizations

Recent research on misleading visualizations has shown that deception in charts often arises through scale manipulation, truncation, salience bias, inappropriate area encoding, dual axes, framing labels, and other design tactics that alter interpretation without necessarily falsifying data (Camba et al., 2022; Lauer and O'Brien, 2020; Lisnic et al., 2023; Driessen et al., 2022). Experimental evidence suggests that some misleading chart formats, particularly truncated bar graphs, can produce persistent interpretive distortions even after viewers receive warnings or basic training. Recent educational reviews have further emphasized that misleadingness in visualizations is not solely a design problem, but also a pedagogical one (Rho and Rau, 2025).

A particularly important step in this direction is the work of Rho et al. (2026), who compared the deceptive impact of different misleading visualizations and discussed implications for adaptive data-literacy support in computer-based learning environments. The literature suggests, first, that common misleading techniques can be meaningfully grouped into recurrent visual families, such as axis manipulations, scale distortions, dual-axis constructions, and

Table 1
Condensed retrospective trajectory of the digital twin concept toward multimodal cognitive digital twins of learners.

Theme / stage	Main contribution to the trajectory	Ref. Year
Foundational cognitive and perceptual background		
Framing and decision making	Established that equivalent information can produce different decisions depending on formulation, providing a cognitive basis for later work on textual misleadingness.	(Tversky and Kahneman, 1981)
Perceptual context effects	Linked size-contrast illusions to information-processing mechanisms, supporting the claim that visual interpretation is context-dependent.	(Gold, 2014)
Dynamic visual illusion effects	Demonstrated that contextual size illusions can be amplified under dynamic conditions, reinforcing the relevance of perceptual instability for multimodal interpretation.	(Mruczek et al., 2015), (Guess et al., 2020).
Visual illusion synthesis	Provided a broad scholarly synthesis of geometric-optical illusions, establishing a theoretical foundation for context-driven misinterpretation in visual materials.	(Shapiro and Todorovic, 2017)
Ebbinghaus reinterpretation	Refined the interpretation of the Ebbinghaus illusion as a size-contrast phenomenon, further clarifying how contextual visual structure distorts judgment.	(Todorović and Jovanović, 2018)
Misleadingness in charts, texts, and multimodal interpretation		
Misinformation resilience in education	Showed that digital media literacy interventions can improve discernment between reliable and false information, establishing an important educational intervention baseline.	(Guess et al., 2020)
Visual theory of mind for misleading charts	Introduced CHARTOM, a benchmark that separates factual chart understanding from the ability to predict how humans may be misled by visualizations.	(Bharti et al., 2024)
Narrative theory of mind in long context	Showed that theory-of-mind reasoning in textual settings depends on contextual understanding over extended narrative structure, relevant for text-based misleadingness.	(Zhou et al., 2025)
Educational approaches to misleading visualizations	Reviewed how misleading visualizations can be addressed pedagogically, shifting the focus from chart design alone to learner support and instructional response.	(Rho and Rau, 2025)
Adaptive support for deceptive visualizations	Compared the deceptive impact of different misleading visualizations and explicitly connected this problem to adaptive data-literacy support in computer-based learning environments.	(Rho et al., 2026)
Digital twins: from industrial systems to education		
Industrial / manufacturing digital twins	Consolidated digital twins as a key Industry 4.0 concept linked to cyber–physical systems, smart manufacturing, and lifecycle integration.	(Tao et al., 2019)
Cognitive digital twins in manufacturing	Introduced cognitive digital twins for smart manufacturing, extending digital twins from monitoring and simulation toward semantic and intelligent decision support.	(Ali et al., 2021)
Formalization of cognitive digital twins	Systematically explored the concept of the cognitive digital twin from a model-based systems engineering perspective and emphasized semantic integration, knowledge representation, and lifecycle reasoning.	(Lu et al., 2022)
Early educational application	Demonstrated one of the early education-oriented uses of digital twins as learning and training support in construction education and campus-based cyber–physical learning environments.	(Dai and Brell-Cokcan, 2022)
Process-level operational twinning	Extended digital twinning toward manufacturing process flows and operational optimization, showing the move from single assets to broader process-level systems.	(Lee and Yang, 2023)
Cognitive multi-agent industrial twins	Advanced cognitive digital twins toward collaborative and intelligent manufacturing systems, including multi-robot and agent-based reasoning.	(Xu et al., 2024)
Education-focused review	Reviewed digital twins in education as an emerging field, indicating that educational use remains relatively recent and less mature than industrial applications.	(Bachmann et al., 2024)
Student-centered digital twins	Proposed AI-based digital twins of students for competency-oriented and personalized learning, shifting the focus from environments to the learner as the core unit of twinning.	(Kabashkin, 2025)

illusion-based graphical effects. Second, controlled experimental research indicates that these families differ substantially in their deceptive impact on interpretation accuracy. Third, learners' data literacy appears to function as a protective factor, but not in a uniform way across all misleading types: some distortions can be mitigated more effectively than others, whereas certain visually or conceptually demanding manipulations may remain highly deceptive even for relatively literate learners. For educational technology, this has an important consequence. If misleadingness varies by visualization type and learner profile, then computer-based support should not be generic, but diagnostically informed and adaptive. This insight directly motivates the present study's move toward learner-centered digital twins and integral indices that can represent not only whether learners are misled, but also by which type of manipulation and with what degree of vulnerability.

2.3. Textual framing, suggestive wording, and false interpretation

A parallel body of literature shows that misleadingness also arises through language. Research on framing effects demonstrated that equivalent information can lead to systematically different decisions depending on how it is described (Tversky and Kahneman, 1981). Likewise, research on language and memory showed that suggestive wording can alter recollection, interpretation, and confidence (Loftus and Palmer, 1974). In contemporary digital environments, textual misleadingness often takes the form of selective omission, confidence-laden wording, causal insinuation, perspective narrowing, and emotionally loaded framing.

These mechanisms are particularly important in a school setting because students are increasingly exposed to information through various channels, including TikTok videos, short online texts, captions, prompts, headlines and platform-based instructional materials, as well as traditional textbook-style explanations. In such contexts, misleading information may not only come from outright falsehoods, but also from the way information is framed, what is omitted and the interpretation implicitly encouraged.

For this reason, the present study does not use neurolinguistic programming as a theoretical construct in the popular sense. Instead, it draws on more well-established research traditions concerning framing, suggestive language and the effects of misinformation (Sturt et al., 2012). This perspective is consistent with the present article's broader goal of modeling how learners in school-based environments may be misled not only by charts, but also by textual formulations and by the interaction between verbal and visual cues.

2.4. From separate benchmarks to multimodal misleadingness

Although misleading charts and misleading texts have both been studied, they are often treated as separate research problems. This separation is increasingly inadequate in educational contexts, where learners frequently encounter mixed materials in which graphical and textual cues interact. Recent benchmark development reflects the emergence of this broader issue. CHARTOM introduced a visual theory-of-mind benchmark for multimodal large language models (Bharti et al., 2024). On the textual side, CharToM-related work has shown that successful theory-of-mind reasoning depends on contextual understanding in extended narrative environments (Zhou et al., 2025). However, these two lines of work remain largely disconnected.

As shown in Table 1, index- and scale-based approaches have been demonstrated to be useful in related fields, such as measuring misleading charts and susceptibility to misinformation and disinformation. These studies suggest that composite measurement frameworks can capture complex forms of interpretive vulnerability more adequately than single observed variables alone. Building on this line of work, the present study extends the index-based approach to a school-based, multimodal setting. This is achieved by integrating misleadingness in charts, text, and mixed charts and text into a unified learner-level modelling framework centred on the HMMI. Index- and scale-based approaches are an effective way to operationalize complex forms of cognitive and informational vulnerability. In the visual domain, Bharti et al. (2024) introduced a human misleadingness index logic for misleading charts within a visual theory-of-mind benchmark. In misinformation research, Maertens et al. (2024) developed the Misinformation Susceptibility Test (MIST) as a psychometrically validated composite measure of news-veracity discernment, while Jin et al. (2021) proposed a composite disinformation susceptibility index at the county level, and Katsiroumpa et al. (2025) validated an online misinformation susceptibility scale. These studies confirm the methodological value of composite indices for representing multidimensional susceptibility. However, they remain either domain-specific or scale-based and do not address school-based multimodal misleadingness in an integrated way. Building on this line of work, the present study introduces the HMMI as an integral index of multiplicative form, designed to quantify learner vulnerability across charts, texts, audio, video, and mixed materials. Unlike simple additive summaries, our index is intended to capture the interaction of error, confidence, persistence, and modality-specific distortions within one interpretable diagnostic measure (An increase in the index in our study indicates that pupils are better prepared).

A systematic review of the scientific literature (Table 1) is key to developing the concept of the digital twin as a promising foundation for the present framework because it allows for the integration of diverse learner data into an interactive, educational representation. In educational research, digital twins are computational representations of learners, learning processes, or instructional environments that support diagnosis, prediction, and intervention design (Bachmann et al., 2024; Kabashkin, 2025).

The present study adopts a more cognitively grounded interpretation. A useful learner digital twin should not merely estimate whether a learner will succeed or fail. Rather, it should model how a learner is likely to be misled, why a misunderstanding emerges, how confidently it is maintained, and what kind of support is most likely to correct it. In this sense, the digital twin is conceived not only as a predictive model, but also as a diagnostic and intervention-design instrument.

From a computational perspective, such learner-centered digital twins require an integrated software ecosystem rather than a single model. In the present framework, the most relevant components include a learning management system for behavioral and assessment traces, annotation tools for multimodal learner data, computer vision and handwriting-recognition pipelines for chart and response processing, large language models for reasoning-trace interpretation and adaptive feedback generation, vector databases for learner-memory representations, and experiment-tracking tools for reproducible model development. Concretely, this architecture can be implemented through tools such as Moodle, Python-based orchestration, Label Studio, YOLO-based document and chart segmentation, OCR/HTR modules, LLM-based reasoning analysis, Qdrant for semantic memory, and MLflow for experimental tracking.

This perspective aligns with the growing body of educational work on misinformation resilience and digital literacy that emphasizes the need for targeted interventions rather than generic awareness alone (Guess et al., 2020; Ali and Qazi, 2023). In the context of the present study, the digital twin therefore functions as the computational core that connects multimodal learner evidence with adaptive support decisions.

Overall, the literature suggests that existing research has provided valuable insights into misleading visualizations, textual framing, and learner support. However, these insights have rarely been combined within a single, actionable educational model. This study proposes an integrated framework that treats charts, texts, and mixed chart-text materials as interconnected modalities of misleadingness rather than isolated phenomena. Based on this perspective, we introduce a methodology combining multimodal cognitive digital twins and the HMMI to diagnose, compare, and reduce learner vulnerability across modalities. Based on the theoretical background and literature review outlined earlier, the next section will define the study's research aims, questions, and hypotheses.

3. Research aims, research questions, and hypotheses

3.1. Research aim

The aim of this study is to empirically evaluate how different types of misleading school-based visualizations affect learners' interpretation accuracy and how these effects vary as a function of learners' data literacy. More specifically, the study builds on an existing prototype of a learner-oriented cognitive digital twin implemented in an open-source Moodle environment and extended with integrated AI functionality (AIPLPGPT AI Powered Learning Platform GPT 5). Within this focused scope, the study examines whether the prototype can serve as an effective school-based technology for diagnosing and reducing learner vulnerability to misleading visual materials.

In the proposed framework, the cognitive digital twin is treated not merely as a performance tracker, but as a learner-centered diagnostic model linked to core cognitive functions involved in interpretation, including perception, attention, memory, reasoning, confidence monitoring, and error correction. Accordingly, the framework is designed to capture not only whether a learner answers correctly or incorrectly, but also which misleading visual features most strongly induce error, how confidently incorrect interpretations are maintained, and to what extent they can be corrected through targeted support.

3.2. Research objectives

To achieve this aim, the study develops a school-based and educationally actionable framework for modeling learner vulnerability to misleading visualizations through a cognitive digital twin architecture. Within this framework, the HMMI is operationalized as a composite measure of learner vulnerability that integrates interpretation error, confidence-weighted error, persistence after feedback, and related cognitive-response indicators. Empirically, the study examines how different misleading visualization types vary in their impact on learner performance, how these effects are moderated by learners' data literacy, and how effectively the AIPLPGPT prototype can identify misleadingness profiles and support adaptive intervention in school-based learning.

3.3. Research questions

The study is guided by the following research questions:

RQ1. In what ways do different types of misleading school-based visualisations impact learners' interpretation accuracy and related cognitive response patterns?

RQ2. To what extent does learners' data literacy moderate the impact of misleading visualizations on interpretation accuracy, confidence, and persistence of error?

RQ3. To what extent does the HMMI provide a measure of learner understanding and resistance to misleading visual materials that is both interpretable and educationally useful?

RQ4. To what extent does the AIPLPGPT cognitive digital twin prototype effectively identify misleading learner profiles and support adaptive intervention in school-based tasks?

3.4. Hypotheses

Based on prior work on misleading visualizations, data literacy, and adaptive instructional support, the following confirmatory hypotheses are formulated:

H1. Different misleading visualization types will differ significantly in their impact on learners' interpretation accuracy, with some visual manipulations producing stronger cognitive distortion than others.

H2. Learners with lower data literacy will be more vulnerable to misleading visualizations, showing lower accuracy, higher confidence in incorrect interpretations, and greater persistence of error.

H3. The HMMI will differentiate learner vulnerability more effectively than binary accuracy alone by capturing the joint effect of error, confidence, and correction resistance.

H4. The AIPLPGPT cognitive digital twin prototype will support more effective diagnosis of misleadingness profiles and greater reduction of posttest vulnerability than non-adaptive support.

3.5. Exploratory dimension

In addition to the confirmatory hypotheses, the study includes an exploratory component concerning individual differences in cognitive vulnerability to misleading visualizations. In particular, the analyses explore whether variables such as school level, age, prior graph literacy, reading confidence, subject background, and gender are associated with different misleadingness profiles.

4. Methods

4.1. Research design

The study was designed as a fully online school-based intervention implemented in an open-source Moodle environment and extended with integrated AI functionality through the AIPLPGPT prototype. The design was aligned with the central purpose of the study: to examine how different forms of feedback and learner support influence students' vulnerability to misleading visual, textual, and mixed chart–text materials.

The study followed a pretest–training–posttest structure with three instructional conditions. In the first condition, learners worked with a full digital-twin-supported environment. In this condition, student errors were not treated in a purely binary way as simply correct or incorrect. Instead, after an incorrect response, the system generated individualized explanatory support and small stepwise hints adapted to the learner's error profile. Typical prompts included short guidance such as using a ruler, re-checking scale intervals, comparing labels more carefully, or revisiting the visual structure of the task.

In the second condition, learners worked with a partial digital-twin-supported environment. Here, digital-twin functionality was used only in a limited or incomplete manner, so that learners received some adaptive support, but not the full range of individualized error-sensitive guidance available in the first condition.

In the third condition, learners followed a standard feedback format. In this condition, tasks were completed in a more traditional way, and feedback remained binary, indicating only whether the response was correct or incorrect, without individualized explanatory prompts. A central methodological contribution of the present work is that these findings became accessible through the use of handwritten-text recognition and OCR-supported analysis, which made it possible to move beyond binary scoring and toward the cognitive interpretation of authentic written learner responses.

Importantly, the pretest phase refers to the stage after initial teacher-led instruction, when all groups had already received the relevant topic teaching and then moved to independent test completion. The posttest phase refers to measurement after the subsequent learning period, during which the three groups differed in the type of feedback and digital-twin support they received. Thus, pretest values represent baseline learner vulnerability after common teacher instruction, whereas posttest values reflect learner performance after exposure to condition-specific support.

4.2. Participants

The analytic sample consists of 280 students from an international gymnasium. Of these, 60% belong to the middle-school group (Grades 5–8) and 40% to the upper-secondary group (Grades 9–11). All participation was online. For the present analysis, the learner is the primary unit of analysis. For clarity, the three instructional conditions are interpreted substantively in the present article as a full digital-twin condition, a partial digital-twin condition, and a standard-feedback condition. In the underlying dataset, these conditions are represented by technical group labels used for coding and analysis.

4.3. Materials and procedure

The study focused on misleading school-based tasks involving visual, textual, and mixed chart–text materials. Task materials were organized so that learners had to interpret information under conditions that could induce perceptual, conceptual, or framing-related error. Across the platform, tasks were implemented in formats that allowed the collection of both response outcomes and learner-support traces.

The instructional procedure differed by condition. After the common teacher-led instructional phase, all learners completed the pretest independently. This pretest therefore captured the learners' initial level of vulnerability after shared classroom teaching and before condition-specific support was applied.

During the subsequent training phase, the three groups received different forms of feedback. In the full digital-twin condition, each incorrect learner response triggered individualized support rather than only binary evaluation. The system provided short adaptive explanations and micro-hints tailored to the likely source of error. These prompts were designed to support self-correction and could include guidance such as checking a visual scale again, measuring a segment with a ruler, comparing proportions more carefully, or revisiting the wording of the task.

In the partial digital-twin condition, learners received a more limited form of such support. Some elements of adaptive guidance were present, but the intervention was less individualized and less comprehensive than in the full digital-twin condition.

In the standard-feedback condition, learners completed the same or equivalent task material under conventional test-like feedback. Responses were marked only in binary form as correct or incorrect, without individualized follow-up explanation or adaptive hints.

After this learning phase, all learners completed the posttest. The posttest was used to evaluate how the different forms of feedback and digital-twin support were associated with subsequent changes in learner-level misleadingness as captured by the modality-specific subindices and the overall HMMI.

To broaden the educational relevance of the design, some tasks are embedded in authentic school-like contexts, such as mathematics, science, media-literacy, or social-studies interpretation tasks. Across the platform, tasks are implemented in two complementary response modes: a conventional closed-response mode scored as correct or incorrect, and an AI-supported mode in which learner explanations, confidence statements, and correction traces are additionally analyzed. Each task therefore contains at least three components: an interpretation response, a confidence judgment, and a corrective follow-up step for persistence coding.

In the analytical dataset, the instructional conditions are coded as follows: 1) Full digital twin condition; 2) Partial digital twin condition; 3) Standard feedback condition. The full digital twin condition (FDT) provides personalised, non-binary, error-sensitive feedback after incorrect responses, including short, adaptive explanations and micro-hints tailored to the learner's likely source of error. The partial digital twin (PDT) condition provided only a limited form of this support. The standard feedback (ST) condition relied on conventional binary feedback, indicating only whether a response was correct or incorrect.

4.4. Data structure and variables

The central analytical dataset is organized at the learner–task level. Each row corresponds to one learner's response to one task and contains the minimum set of variables required for subsequent statistical analysis, HMMI construction, and adaptive profile modeling. The final dataset used in the present analysis contains 280 learner-level records, with one row per learner. This structure makes it possible to aggregate results upward to the task level, the misleadingness-type level, the learner level, the school-level group, and the intervention level.

A6 (accuracy) captures the correctness of the learner's interpretation in dichotomous form. A7 (deviation magnitude) represents the normalized distance between the learner's response and the expected interpretation. A8 (confidence) reflects the learner's self-reported certainty. A9 (persistence) indicates whether an incorrect interpretation remains after minimal corrective prompting. A10 (data literacy) is treated as a learner-level covariate derived from a separate literacy measure or pretest score. A11 (response time) serves as an auxiliary behavioral indicator of cognitive processing load. A12 contains the task-level HMMI value derived from the observed learner response.

4.5. Taxonomy of misleadingness types

Compared with prior studies that focus primarily on misleading visualizations alone, the present framework covers a broader spectrum of misleading school tasks. In addition to classical visual manipulations such as truncated axes, dual axes, 3D effects, and inverted scales, the design also includes textual misleadingness, mixed text–image misleadingness, and cognitively overloaded tasks in which the essential information is embedded within distracting material. For consistency with the learner–task dataset, all misleadingness categories are coded under variable A4.

4.6. Digital-twin pipeline

The methodological core of the study is a learner-oriented cognitive digital twin pipeline that transforms platform traces, response data, confidence judgments, and intervention outcomes into learner-specific diagnostic profiles. The school-based learning environment provides learner evidence from interpretation tasks, confidence judgments, and corrective responses. These data are integrated into a learner digital twin and processed through a cognitive learner model. The model supports statistical estimation of HMMI and related subindices, as well as learner-profile interpretation for adaptive instructional support.

4.7. Outcome measures

In contrast to prior studies that assess misleading visualizations mainly through accuracy differences, odds ratios, and moderation by data literacy, the present study adopts an index-based analytic strategy. Rather than treating correctness as the sole outcome, it constructs the HMMI as an integral measure combining error, deviation, confidence, persistence after feedback, and related learner-level indicators. In addition, unlike conventional designs based primarily on closed-response items, the present study includes open handwritten responses. Their analysis provides a broader diagnostic spectrum, including traces of reasoning, structural misconceptions, partial understanding, symbolic errors, and correction dynamics.

4.8. Measures and HMMI construction

At the task level, the basic observed components are defined for learner i on task t as follows:

$$E_{it} \in [0,1] \quad \text{error occurrence,} \tag{1}$$

$$D_{it} \in [0,1] \quad \text{deviation magnitude,} \tag{2}$$

$$C_{it} \in [0,1] \quad \text{confidence in the selected interpretation,} \tag{3}$$

$$P_{it} \in [0,1] \quad \text{persistence after corrective prompting.} \tag{4}$$

All indicators are normalized to the interval $[0, 1]$ using min–max normalization:

$$z_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}}. \tag{5}$$

Table 2
Excerpt from the core learner–task dataset used for statistical analysis and HMMI construction.

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
S001	5–8	FDT	A4.1.1.1	MCQ	0	0.72	0.84	1	0.41	19.2	0.74
S002	5–8	ST	A4.1.3.1	MCQ	1	0.18	0.46	0	0.52	15.6	0.28
S003	5–8	PDT	A4.1.2.2	Open	0	0.61	0.79	1	0.39	21.4	0.68
S004	9–11	FDT	A4.1.1.6	MCQ	1	0.11	0.44	0	0.73	13.8	0.79
S005	9–11	ST	A4.1.4.1	Open	0	0.49	0.63	1	0.66	17.1	0.53
S006	9–11	PDT	A4.1.1.2	MCQ	1	0.09	0.38	0	0.77	12.4	0.15
S007	5–8	FDT	A4.2.1.1	Open	0	0.67	0.81	1	0.35	22.6	0.71
S008	9–11	PDT	A4.1.2.5	MCQ	1	0.14	0.42	0	0.81	11.9	0.17

Note. A1 = learner ID (S001-S280); A2 = grade group; A3 = instructional condition; A4 = misleadingness type code; A5 = task format; A6 = accuracy (1 = correct, 0 = incorrect); A7 = deviation magnitude; A8 = confidence; A9 = persistence after corrective prompting; A10 = data literacy score; A11 = response time in seconds; A12 = task-level HMMI.

Table 3
Hierarchical coding of misleadingness types under variable A4.

Code	Category	Example task form
A4.1	Visual misleadingness	Visual distortions in charts, diagrams, and graphical representations
A4.1.1	Axis and scale manipulations	Truncated axis, compressed axis, inverted axis, manipulated spacing
A4.1.2	Form and salience distortions	3D effects, pictorial bars, pop-out emphasis, visual salience
A4.1.3	Relation distortions	Dual axes, false visual correlation, hidden proportional mismatch
A4.1.4	Geometric-perceptual misleadingness	Ebbinghaus-type size contrast, distorted coordinate reading
A4.2	Textual misleadingness	Wording, framing, omission, and question-formulation traps
A4.2.1	Framing and wording effects	Misleading caption, suggestive headline, biased comparison prompt
A4.2.2	Context omission	Selective omission, partial explanatory context, irrelevant persuasive detail
A4.2.3	Question traps	Hidden negation, false assumption, ambiguous wording
A4.3	Mixed text-image misleadingness	Tasks where textual and visual distortions reinforce one another
A4.3.1	Text reinforces visual distortion	Chart + biased caption, diagram + misleading explanation
A4.3.2	Mixed school media tasks	Science figure + misleading label, geography map + biased legend
A4.4	Cognitive overload and trap tasks	Tasks with informational noise, hidden core, or conceptual traps
A4.4.1	Cognitive overload	Long A4 narrative, excessive irrelevant detail, high reading load
A4.4.2	Mathematical and logical traps	Pizza radius task, fake proportionality, unit-conversion trap
A4.4.3	Hidden-core tasks	Core question embedded in narrative context, delayed discovery of variable

For indicators for which larger values correspond to lower misleadingness, inverse normalization is applied:

$$z_{ij} = \frac{x_j^{\max} - x_{ij}}{x_j^{\max} - x_j^{\min}}. \tag{6}$$

For each learner i and task t, task-level misleadingness is modeled in multiplicative form:

$$\mathrm{I}_{it} = E_{it}^{\alpha_E} \cdot D_{it}^{\alpha_D} \cdot C_{it}^{\alpha_C} \cdot P_{it}^{\alpha_P}, \tag{7}$$

where α_E, α_D, α_C, and α_P are theoretically specified or empirically calibrated component weights. In the baseline specification used in the present analyses, equal weights are assumed.

Learner-level HMMI is obtained by aggregating task-level values into broader misleadingness groups and then combining the resulting group-specific subindices:

$$\mathrm{HMMI}_i = \prod_{g=1}^{G} \left(I_i^{(g)} \right)^{\beta_g}, \tag{8}$$

subject to

$$\sum_{g=1}^{G} \beta_g = 1, \tag{9}$$

where $I_i^{(g)}$ denotes the subindex for misleadingness group g, and β_g denotes the corresponding group weight. This multiplicative form is preferred because it is sensitive to the joint effect of high-risk components.

In the present study, the learner-level HMMI is operationalized through three modality-specific subindices:

$$I_j \in [0, 1], \qquad j = 1, 2, 3,$$

where I_1 denotes the visual misleadingness subindex, I_2 the textual misleadingness subindex, and I_3 the mixed chart–text misleadingness subindex.

To avoid degenerate zero values in the multiplicative index, all normalized components are shifted by a small positive constant $\varepsilon > 0$ before aggregation. Thus, instead of using the raw normalized values in $[0, 1]$, the empirical implementation applies an ε-adjusted scale so that each component lies in $[\varepsilon, 1]$. This preserves the ordinal meaning of the indicators while ensuring that a single zero-valued component does not collapse the entire product to zero. Accordingly, the task-level index reflects relative learner vulnerability even in cases where one component would otherwise take the boundary value 0.

This formulation preserves the multiplicative logic of the index while making explicit that overall learner vulnerability is jointly determined by visual, textual, and mixed chart-text misleadingness. In the empirical implementation used in this study, these three subindices are represented as learner-level measures and then combined into the overall HMMI, which is calculated at both pretest and posttest, that is, before and after the condition-specific learning phase.

4.9. Statistical analysis

The statistical analysis follows a two-layered strategy aligned with the study's confirmatory aims. First, conventional performance-based modeling is used to estimate how misleading versions of school-based tasks affect the probability of correct interpretation relative to their non-misleading counterparts. Second, an index-based analysis is used to quantify learner understanding more broadly through the HMMI, which is computed as a multiplicative composite index from task-level components and then aggregated across the three modality-specific learner-level subindices defined above.

For multiple pairwise comparisons, Holm-adjusted p-values were used to control the family-wise error rate. Let $p_{(1)} \le p_{(2)} \le \cdots \le p_{(m)}$ denote the ordered raw p-values for a family of m hypotheses. Under the Holm procedure, hypothesis $H_{(k)}$ is rejected if

$$p_{(k)} \le \frac{\alpha}{m - k + 1}, \qquad k = 1, \ldots, m, \tag{10}$$

in a step-down sequence until the first non-rejected hypothesis is reached.

To test differences in interpretation accuracy in the general analytical framework of the study, a generalized linear mixed-effects model (GLMM) with a logit link is specified for the binary accuracy outcome. Fixed effects include task version (misleading vs. non-misleading), misleadingness type, data literacy, instructional condition, and their theoretically motivated interactions. Random intercepts for learners and items are included to account for repeated measures and item-level variability. Odds ratios (ORs), 95% confidence intervals, and estimated marginal means are reported. Post hoc contrasts and slope comparisons are obtained through estimated marginal means analysis.

To evaluate learner vulnerability beyond binary correctness, task-level HMMI is first computed and then aggregated to the learner level. In the empirical dataset used in the present study, learner-level pretest and posttest HMMI values are analyzed descriptively and comparatively across instructional conditions and grade groups. In addition, posttest HMMI is modeled as a function of pretest HMMI, instructional condition, and learner-level background characteristics in order to evaluate whether adaptive digital-twin-guided support is associated with lower posttest vulnerability than generic or non-adaptive support.

Descriptive statistics are reported for the main learner-level variables, including the modality-specific subindices and the overall HMMI. All analyses are implemented in R. Mixed-effects models are fitted with `lme4`; estimated marginal means, contrasts, and slope comparisons are computed with `emmeans`; and publication-ready figures are produced with `ggplot2`. The significance level is set at $\alpha = .05$. Model estimates are reported together with standard errors, confidence intervals, and effect-size-relevant quantities.

Supplementary R scripts, extended figures, and additional output tables will be made available in an online project repository upon publication.

5. Results

5.1. Sample overview and baseline comparability

The final analytic sample consisted of 280 learners. The instructional conditions comprised a full digital-twin condition ($n = 131$), a partial digital-twin condition ($n = 110$), and a standard-feedback condition ($n = 39$). The sample included learners from two grade-group categories, Grades 5–8 and Grades 9–11.

To examine baseline comparability across the three instructional conditions, pretest HMMI scores were compared before the differentiated feedback phase. No statistically significant between-group difference was observed for pretest HMMI, indicating that the three groups were broadly comparable at baseline prior to the condition-specific intervention phase.

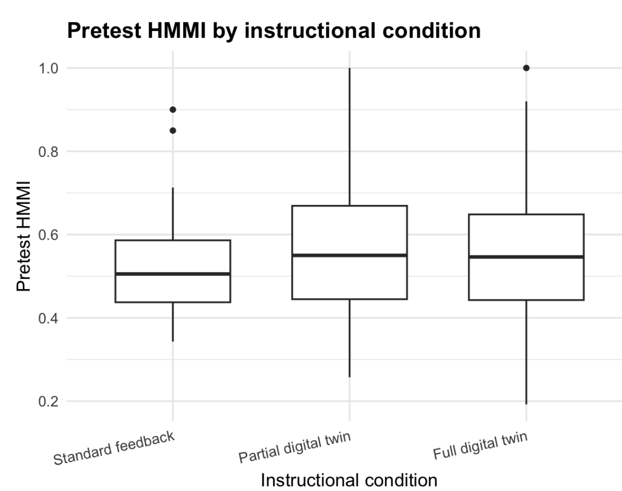

Fig. 1. Distribution of pretest HMMI across the three instructional conditions.

Pretest HMMI by instructional condition. After a common instructional phase devoted to a new topic on working with graphs, learners were randomly divided into three groups and completed a shared pretest before the intervention phase. In the full digital-twin condition, student errors were not treated in a purely binary way; instead, each incorrect response triggered individualized support in the form of short explanations and micro-hints tailored to the likely source of error, such as using a ruler, re-checking scale intervals, or comparing labels more carefully. Descriptively, the full digital-twin group appears to include learners with somewhat weaker initial performance, although the pretest between-group differences were not statistically significant.

5.2. Descriptive structure of the three modality-specific subindices

Having established the baseline structure of the three instructional groups, the next step was to examine the internal composition of learner-level multimodal misleadingness at pretest. For this purpose, we analyzed the three modality-specific subindices that constituted the overall HMMI, namely the visual misleadingness subindex (I_1), the textual misleadingness subindex (I_2), and the mixed chart–text misleadingness subindex (I_3).

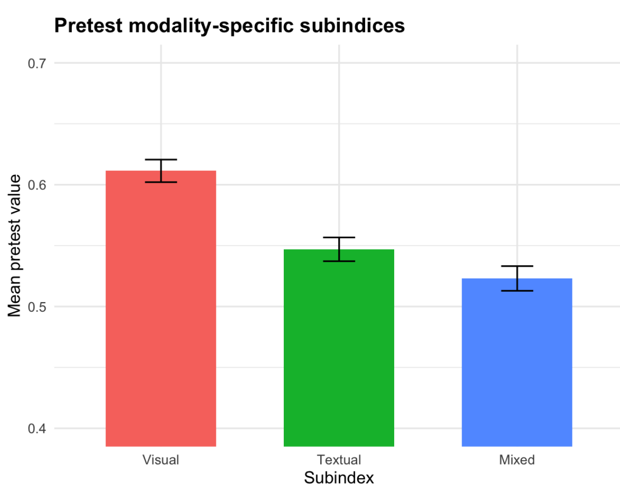

Fig. 2. Mean pretest values of the visual, textual, and mixed misleadingness subindices across the full learner sample.

The visual subindex was derived from a set of graph- and visualization-based tasks representing recurrent misleadingness families. These included line charts with truncated y-axes, compressed y-axes, inverted y-axes, and manipulated x-axes; line charts with dual axes; scatter plots with logarithmic y-axes and flipped x-axes; bar charts with truncated or compressed y-axes, pictorial bars, and 3D effects; pie charts with 3D effects and pop-out emphasis; and map-based tasks with inverted color scales. The visual-illusion-related component of the present results was highly consistent with the pattern reported by Rho and colleagues (Rho et al., 2026). In contrast to that study, the present framework extended the visual task set by incorporating classical perceptual illusions, especially Müller–Lyer-type items (Fig. 3), which emerged as some of the most difficult tasks for learners in our sample.

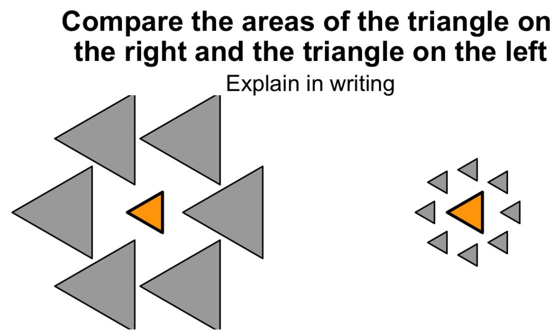

Fig. 3. Example of a Müller–Lyer-type visual task included in the visual misleadingness component of the study.

In addition to graph-based distortions, the visual task pool was expanded by including classical perceptual illusions. Two representative examples are shown in Figs. 4 and 5. The first type involved tile-based parallel-line illusions related to the Café Wall family, in which objectively parallel horizontal boundaries may appear slanted or non-parallel because of local contrast and offset structure. The second type involved Müller–Lyer-type items, in which equal line segments may be perceived as different in length because of contextual arrow-like cues (Shapiro and Todorovic, 2017; Todorović and Jovanović, 2018).

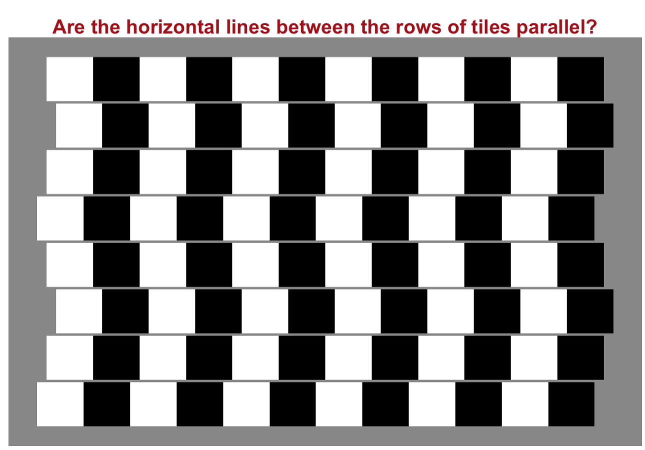

Fig. 4. Example of a tile-based parallel-line illusion related to the Café Wall family.

Although the horizontal boundaries are objectively parallel, their arrangement may induce the false impression that they are tilted or non-parallel.

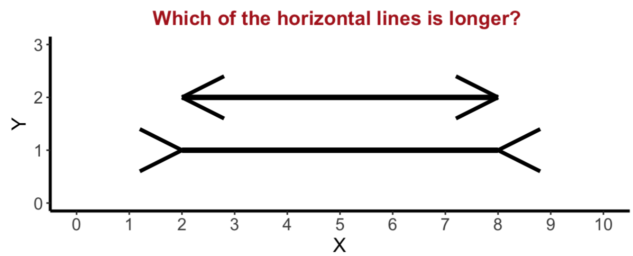

Fig. 5. Example of a perceptual parallel-line illusion task.

The two horizontal line segments are equal in length, but the surrounding arrow-like cues may bias visual comparison and create the impression that one segment is longer than the other. Such items were especially useful for identifying learners who relied primarily on immediate perceptual impression rather than explicit measurement or analytic comparison.

The figure illustrates the internal structure of the overall HMMI at baseline, prior to the consideration of condition-specific pretest-posttest comparisons. This analysis is important because the HMMI is an aggregated learner-level measure. Examining its three components separately enables us to determine whether the initial pattern of learner vulnerability was concentrated primarily in one modality or distributed more broadly across visual, textual and mixed tasks. Accordingly, the present results not only replicate the broader finding that misleading visual formats differ in deceptive power, but also extend it by demonstrating that illusion-based perceptual distortions constitute a particularly demanding class of school-relevant misleadingness.

The textual subindex was based on tasks in which misleadingness arose primarily through wording, framing, omission, or question formulation. These tasks included both short and extended school-style texts with hidden assumptions, selective contextual omission, suggestive wording, misleading captions, and prompts designed to test whether learners could distinguish between literal recall, inferred meaning, and framed interpretation. The textual pool also included tasks in which essential quantitative information was deliberately absent, so that learners had to recognize that a numerical conclusion could not be validly drawn from the text alone. In addition, some items described geometric, logical, or proportional situations verbally rather than visually, requiring students to construct an internal representation of the problem instead of relying on a given diagram. This category also included underdetermined tasks, overdetermined tasks, and tasks with internally inconsistent premises, that is, problems that had no valid solution, more than one possible interpretation, or insufficient information for a unique answer. Such formats were included to test whether learners would critically evaluate the adequacy of the textual information or instead proceed automatically toward a premature but incorrect answer. More broadly, the textual subindex was intended to capture not only reading comprehension difficulties, but also learners' vulnerability to memory load, implicit assumptions, incomplete specification, and linguistically induced false certainty.

The mixed chart–text subindex was constructed from tasks in which visual and verbal cues interacted. In these tasks, learners had to interpret a graph, diagram, or chart together with a caption, explanatory note, headline, legend, or short contextual paragraph. This format was especially important because it reflects realistic school and digital-media situations in which misleadingness does not come from one modality alone, but from the combined effect of visual representation and textual framing. The mixed task pool therefore included items in which a graph was formally readable, but its accompanying text encouraged an incorrect interpretation; items in which the text omitted a crucial condition needed to interpret the visual correctly; and items in which the learner had to reconcile partially conflicting visual and verbal cues. It also included verbally formulated geometry or mathematics tasks accompanied by incomplete, misleading, or visually suggestive diagrams, as well as chart-based items whose captions or explanatory notes encouraged overgeneralization, false comparison, or unsupported causal inference. Across the pretest observations, the combination of textual misleadingness and visual misleadingness proved to be particularly demanding for learners. In many cases, this mixed format appeared to be more difficult than purely visual or purely textual tasks, presumably because students had to coordinate perceptual interpretation, reading comprehension, working memory, and critical evaluation at the same time. In this sense, the mixed chart-text subindex captured not only multimodal misleadingness as such, but also the additional cognitive burden created when two potentially misleading channels reinforce one another. Table **??** summarizes the main textual and mixed task formats used to operationalize the corresponding modality-specific subindices.

In addition to correctness-related indicators, response time was monitored by default for each modality-specific subindex separately. This made it possible to compare not only the level of learner vulnerability across the three components, but also the behavioral effort associated with different misleadingness families. Exploratory inspection of the pretest data suggested that one important source of learner difficulty was not limited to visual interpretation alone. In particular, many children appeared to experience substantial difficulty when reading longer texts and retaining previously presented written information, especially in tasks requiring the integration of several textual cues across a longer prompt. This observation is consistent with the broader motivation of the present study, namely that multimodal misleadingness in school settings often depends on the interaction of perceptual, linguistic, and memory-related demands.

Taken together, the textual and mixed subindices are especially important from the perspective of OCR- and handwriting-based analysis. Unlike purely closed-response formats, these task families provide access to authentic written learner traces, including omissions, unstable interpretations, incomplete justifications, and revision attempts.

Table 4
Representative textual and mixed task formats used in the study.

Subtype	Typical task format
Framing / omission	Short or extended school-style texts with hidden assumptions, selective omission, suggestive wording, misleading captions, and biased prompts.
Underspecified / unsolvable	Word problems without sufficient numerical information, verbally described geometry tasks without diagrams, and inconsistent or underdetermined problem statements.
Memory-load integration	Longer texts requiring retention of several written conditions or integration of multiple cues across the prompt.
Text reinforces visual error	Graphs or diagrams accompanied by captions, legends, headlines, or explanatory notes that encourage an incorrect interpretation.
Visual–text conflict	Tasks in which the graph is readable, but the accompanying text omits a key condition, introduces a false inference, or partially conflicts with the visual.
School-like multimodal tasks	Mathematics, science, or geography tasks combining verbal description with incomplete, suggestive, or misleading diagrams, tables, or charts.

In this sense, the present results represent a first step made possible by the use of handwritten-text recognition, which opens new opportunities for analyzing misleadingness not only at the level of final answers, but also at the level of cognitive response structure and learner-specific misconception patterns.

Across the sample, the three subindices showed comparable mid-range values, indicating that learner vulnerability was not concentrated in only one modality. It is important to note, however, that each modality-specific subindex was constructed from a heterogeneous pool of approximately ten different misleadingness families or task formats rather than from only one or two illusion types. Accordingly, the resulting baseline profile should be understood as a broad learner-level tendency across modalities, not as an artifact of a narrowly defined set of individual items.

In methodological terms, the textual and mixed subindices benefit especially strongly from OCR- and handwriting-based analysis. These formats make it possible to move beyond binary correctness and to examine how learners formulate, omit, revise, and structurally organize their responses in authentic written work. As a result, recent advances in OCR and handwritten-text recognition open particularly important new opportunities for analyzing misleadingness in textual and mixed tasks and for increasing the diagnostic precision of cognitive digital twins (Zalizko, 2025).

5.3. Pretest–posttest change in HMMI across instructional conditions

Having described the internal baseline structure of the three modality-specific subindices, the main analytical step was to examine how learner-level HMMI changed from pretest to posttest across the three instructional conditions. In the empirical coding used in the present study, higher posttest HMMI values corresponded to stronger post-intervention performance on the multimodal misleadingness tasks. The key question was therefore whether the three instructional conditions differed significantly in the magnitude of pretest–posttest change.

Descriptively, the three groups showed clearly different gain patterns. The full digital-twin condition demonstrated the largest increase in HMMI from pretest to posttest, the partial digital-twin condition showed the smallest increase, and the standard-feedback condition occupied an intermediate position. Thus, the observed ordering of gain scores was consistent with the pedagogical expectation that more individualized and non-binary learner support would be associated with stronger post-intervention outcomes.

A one-way analysis of variance conducted on the pretest–posttest change scores confirmed that the three instructional conditions differed significantly in their HMMI gains, $F(2, 277) = 97.66$, $p < .001$. The mean gain in the full digital-twin condition was 0.347, compared with 0.077 in the partial digital-twin condition and 0.199 in the standard-feedback condition. These results indicate that the type of feedback and learner support received during the intervention phase was strongly associated with subsequent changes in HMMI.

Importantly, this condition effect should be interpreted in light of the instructional logic of the experiment. All groups first completed a shared instructional phase and then a common pretest. Only after that point did the groups diverge in the type of support they received. In the full digital-twin condition, incorrect responses triggered individualized explanations and adaptive micro-hints rather than simple binary correctness feedback. In the partial digital-twin condition, such support was available only in a reduced form. In the standard-feedback condition, responses were evaluated conventionally as correct or incorrect without learner-specific adaptive explanation. The resulting gain pattern therefore supports the conclusion that more intensive digital-twin-mediated support was associated with stronger posttest development.

For the main intervention analysis, an individual HMMI gain score was computed for each learner as

$$\Delta\mathrm{HMMI}_i = \mathrm{HMMI}_{i,\mathrm{post}} - \mathrm{HMMI}_{i,\mathrm{pre}}. \tag{11}$$

These learner-level gain scores were then compared across the three instructional conditions using a one-way analysis of variance. Accordingly, the reported group means represent the average pretest–posttest HMMI gain within each instructional condition.

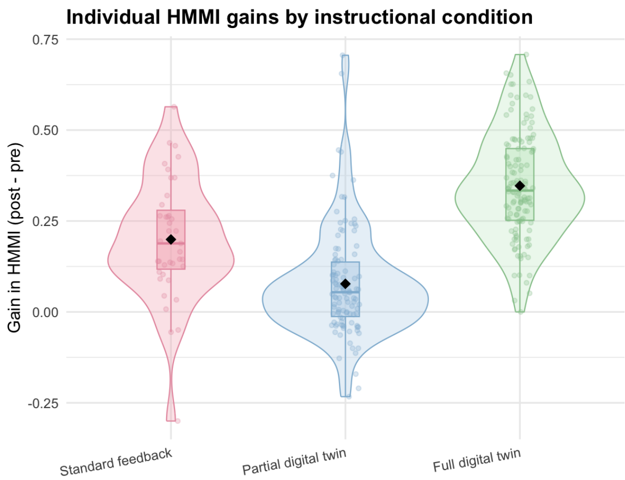

Fig. 6. Mean pretest and posttest HMMI values across the three instructional conditions.

The largest gain was observed in the full digital-twin condition, followed by the standard-feedback condition and the partial digital-twin condition.

To determine whether the observed posttest differences persisted after adjustment for baseline variation, posttest HMMI was additionally modeled as a function of pretest HMMI, instructional condition, grade group, age, and gender. The regression model showed that instructional condition remained a statistically significant predictor of posttest HMMI after controlling for baseline variation, whereas age, grade group, and gender did not emerge as strong independent predictors in the present sample. Thus, the main intervention effect cannot be reduced to simple baseline imbalance or demographic composition alone.

Importantly, the gain pattern shown in Fig. 6 should be interpreted against the shared instructional background of the experiment. All learners were taught the same topic by the same teachers and completed a common pretest before the differentiated support phase began. Under these otherwise comparable instructional conditions, the full digital-twin condition produced the strongest individual gains. This indicates that the most complete form of individualized digital-twin support yielded the highest post-intervention results, above and beyond common teaching input alone.

5.4. Age- and grade-related differences in visual misleadingness

In addition to the main intervention effect, the baseline data revealed an important age-related pattern in the visual component of misleadingness. Specifically, the visual pretest subindex was significantly lower among younger learners (Grades 5–8) than among older learners (Grades 9–11), indicating that visual misleadingness was less readily detected by the younger group. In substantive terms, this suggests that for younger children, deceptive visual structure was less transparent and therefore more difficult to identify critically.

A one-way analysis of variance confirmed a statistically significant grade-group difference for the visual pretest subindex, $F(1, 278) = 52.89$, $p < .001$. The mean visual pretest subindex was 0.561 in Grades 5–8 and 0.687 in Grades 9–11. Thus, although the intervention effect itself was primarily driven by the structure of learner support, baseline sensitivity to visual misleadingness was clearly associated with developmental or school-level differences.

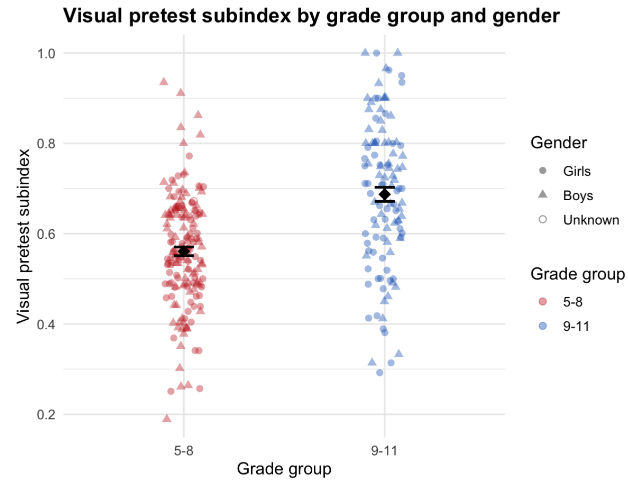

Fig. 7. Visual pretest subindex by grade group and gender.

Each point represents one learner. The age-related difference is more pronounced than any gender-related difference, indicating that misleading visual structure was less readily detected by younger learners, whereas boys and girls showed broadly similar values within each grade group.

This finding is educationally important because it suggests that visual misleadingness cannot be treated as equally obvious across developmental stages. Rather, younger learners appear to require more explicit support in identifying misleading graphical structure, perceptual distortions, and visually persuasive but mathematically incorrect interpretations.

By contrast, gender did not emerge as a strong independent factor in the present analyses. Although small descriptive differences were observed, these did not become statistically prominent once baseline performance and instructional condition were taken into account.

5.5. Summary of the main findings

Overall, the results indicate that, while the three instructional groups were similar at baseline, they diverged substantially by the post-test. The strongest post-test outcomes were associated with the full digital twin condition, followed by the standard feedback condition and then the partial digital twin condition. Additionally, baseline HMMI significantly predicted posttest HMMI. Another important finding was that younger learners showed significantly lower values on the visual pretest than older learners, indicating greater age-related vulnerability to visual misleadingness at baseline. By contrast, gender did not emerge as a strong independent factor in the present analyses. From a methodological perspective, these findings demonstrate that analysing learner responses using OCR and handwriting opens new opportunities for examining misleadingness as a cognitive process rather than only as a final binary outcome.

6. Discussion

The present framework builds upon previous work in several significant ways. It unites misleading charts, textual framing and mixed chart-text materials within a single framework that can be interpreted for educational purposes at learner level. In this sense, the proposed HMMI moves beyond the confines of narrow benchmark logic, enabling multimodal learner vulnerability to be represented in a unified way. The study broadens the scope of previous research on misleading visualisations by shifting the focus from visualisation-centred comparison to learner-centred diagnosis and support. Rather than merely considering which formats are the most deceptive, the framework also considers which learners are most vulnerable and how their vulnerability can be reduced most effectively in each modality and with each type of support. This study should be understood as a first step in this direction, rather than the full realisation of handwriting-driven cognitive twinning. The key contribution at this stage is demonstrating that substantially richer forms of misleadingness analysis become possible once handwritten learner responses can be analysed using OCR-supported methods. This includes identifying not only whether learners were misled, but also how misleadingness was expressed in written reasoning, how stable it remained and what forms of support could reduce it.

Another important practical observation was that many learners found it difficult to produce handwritten explanations for longer text-based tasks. In these cases, the challenges concerned not only correctness, but also omissions, incomplete reasoning, inconsistent formulations and the loss of relevant information in extended written responses. OCR- and handwriting-based analysis proved especially valuable here, as it enabled a systematic review of a large number of learners' work that teachers would not realistically be able to inspect in such depth on a daily basis. In this sense, the digital twin framework does not replace the teacher, but rather expands their diagnostic reach.

The results of the intervention are particularly significant in this respect. All groups received the same topic instruction from the same teachers and completed a common pre-test before the differentiated support phase began. Under these comparable instructional conditions, the full digital twin condition produced the strongest individual gains. This suggests that personalised, non-binary, error-sensitive feedback is more effective in an educational context than reduced adaptive support or conventional correct/incorrect feedback alone. Therefore, the main contribution of the present findings is that learner performance improved substantially depending on the structure of support.

Another important result concerns developmental differences in visual misleadingness. Younger learners showed significantly lower baseline values on the visual pretest subindex than older learners, indicating that deceptive graphical or perceptual structures were less readily detected by the younger group. This implies that visual misleadingness is not equally transparent across age groups, and that younger learners may require more explicit instructional support when working with misleading graphs, perceptual illusions and visually persuasive yet mathematically incorrect representations.

The textual and mixed components of the framework also warrant particular attention. The present study showed that misleadingness in school settings cannot be attributed to visual distortion alone. Learners also struggled with wording, omissions, incomplete specifications, memory load integration and mixed situations in which verbal and visual cues reinforced one another. From a methodological perspective, this is particularly significant because recent analysis based on optical character recognition (OCR) and handwriting opens up new possibilities for examining authentic written learner responses, tracing patterns of misconception, and extending the diagnostic capabilities of learner-oriented cognitive digital twins.

A key curricular implication concerns fake news education. If learners can be misled by multimodal materials with such confidence and persistence, schools should respond with explicit instruction on fake news, misleading graphs, manipulative narratives, source checking and confidence calibration. Therefore, multimodal misinformation resilience should be regarded as a legitimate curricular target, not an optional media literacy add-on.

7. Limitations and future directions

The findings of the present study should be interpreted bearing in mind several limitations. Firstly, although the design was implemented in an authentic, school-based setting, the sample size is limited to one educational context and age range. Future research should extend this framework to a broader range of learner populations, school systems and sociocultural environments.

Secondly, the task pool intentionally included multiple families of misleadingness and heterogeneous response formats. While this broadened the ecological and educational relevance of the study, it also introduced additional variability at the item level. Future work should investigate more systematically how specific task formats, response modes and misleadingness families contribute to differences observed in learner understanding under misleading conditions.

Thirdly, the present study is an early stage in the use of handwritten text recognition for analysing misleadingness. The key methodological advance lies in demonstrating that analysis based on optical character recognition (OCR) and handwriting can provide access to richer cognitive evidence in authentic written learner responses. However, the current framework does not yet realise the full analytical potential of such systems. Future research should develop more refined pipelines for identifying omission patterns, partial reasoning structures, correction attempts and misconception trajectories directly from students' handwritten work.

Fourthly, while the current results suggest that textual and mixed misleadingness benefit particularly from handwritten response analysis, future work should explicitly link these methods to fine-grained learner modelling. In particular, evidence supported by optical character recognition (OCR) could be used to strengthen digital twins by reconstructing evolving reasoning traces, error structures and support-sensitive revision behaviour.

Finally, a key area for future research is extending the current framework to include audio- and video-based misleadingness by adding new, modality-specific sub-indices to the HMMI. In its current empirical form, the index captures visual, textual and mixed chart-text understanding. The next logical step is to incorporate audio and video in order to model a broader range of cognitive vulnerability to misleading cues that unfold over time. This is particularly pertinent in contemporary digital learning environments, where learners are increasingly exposed to misleading spoken explanations, edited short-form videos, and multimodal social media content. Future studies should also investigate how adaptive educational support can be optimised once such multimodal learner evidence is more fully incorporated into the model, including through real-time interventions responsive to individual cognitive profiles.

8. Conclusion

We introduced multimodal cognitive digital twins and the HMMI as a unified educational framework for visual, textual, and mixed misleadingness. In the empirical implementation used in this study, higher HMMI values indicate stronger learner understanding and greater resistance to misleading materials. Using a fully online school-based design with 280 learners, the study showed that the three instructional groups were broadly comparable at baseline, but diverged substantially by posttest. The strongest post-intervention gains were associated with the full digital-twin condition, which achieved a mean HMMI increase of 0.347, compared with 0.199 in the standard-feedback condition and 0.077 in the partial digital-twin condition. These differences were statistically significant, $F(2,277) = 97.66$, $p < .001$, indicating that individualized, non-binary support was associated with the strongest learner improvement.

The study also identified an important developmental pattern in the visual component of misleadingness. Younger learners (Grades 5–8) showed significantly lower baseline visual pretest values than older learners (Grades 9–11), with means of 0.561 and 0.687, respectively, $F(1,278) = 52.89$, $p < .001$. This suggests that misleading visual structure is less transparent to younger learners and may require more explicit instructional scaffolding. By contrast, gender did not emerge as a strong independent factor in the present analyses.

Methodologically, the study demonstrates that analysing learner responses using OCR and handwriting opens up new possibilities for moving beyond binary scoring towards cognitively interpreting authentic written traces of learners, including omissions, partial understanding, revision behaviour and the structure of misconceptions. In this sense, the present work suggests a more diagnostic approach to generating learner-centred digital twins, based on authentic student responses rather than final-answer correctness alone.

An important next step will be to extend the current framework to include audio- and video-based misleadingness, as well as visual, textual, and mixed chart-text materials. This is especially relevant given that many contemporary educational and social media environments present misleading information through spoken explanations, editing, timing, intonation and the structure of multimodal videos rather than static texts or charts.

The present work also highlights an important next step for educational research: many of the most cognitively meaningful aspects of misleadingness only become apparent when authentic handwritten learner responses are analysed rather than reduced to binary correctness alone. In this sense, OCR- and handwriting-based analysis opens up new opportunities for modelling omission patterns, partial understanding, revision behaviour and misconception structure within learner-oriented cognitive digital twins.

More broadly, the findings suggest that misleadingness in school settings is not a marginal phenomenon limited to isolated graph-reading errors. Rather, it emerges through the interaction of perception, framing, omission, memory load and multimodal integration. Therefore, we argue that multimodal misinformation resilience, the critical interpretation of charts and texts, and adaptive learner support based on authentic written learner evidence should be explicit educational priorities in digitally mediated schooling.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, generative AI tools were used to support language drafting and structural editing. The author reviewed and edited the content and takes full responsibility for the final version of the manuscript.

Data availability

Data will be made available on request.

Acknowledgements

This work was developed as an outcome of the author's research activity at the ETH AI Center and collaboration with Professor Gerd Kortemeyer.

CRediT authorship contribution statement

Vasyl Zalizko: Conceptualization, Methodology, Formal analysis, Writing – Original draft, Writing – Review & editing.

References

AI Powered Learning Platform GPT 5. An innovative platform with digital twins of students in grades 1-12. International Innovative School. URL: https://gpt.iis.org.ua/. accessed 27.03.2026.

Ali, A., Qazi, I.A. (2023). Countering misinformation on social media through educational interventions: Evidence from a randomized experiment in Pakistan. Journal of Development Economics 163, 103108. doi:10.1016/j.jdeveco.2023.103108.

Ali, M.I., Patel, P., Breslin, J.G., Harik, R., Sheth, A. (2021). Cognitive digital twins for smart manufacturing. IEEE Intelligent Systems 36, 96–100. doi:10.1109/MIS.2021.3062437.

Bachmann, J.E.C., Silveira, I.F., Martins, V.F. (2024). Digital twins for education: A literature review, in: Anais do XXXV Simpósio Brasileiro de Informática na Educação (SBIE 2024), Sociedade Brasileira de Computação. pp. 722–736. doi:10.5753/sbie.2024.242288.

Bharti, S., Cheng, S., Rho, J., Rau, M., Zhu, X. (2024). CHARTOM: A visual theory-of-mind benchmark for large language models on misleading charts. doi:10.48550/arXiv.2408.14419, arXiv:2408.14419.

Camba, J.D., Company, P., Byrd, V., Sousa Santos, B., Alford, G. (2022). Identifying deception as a critical component of visualization literacy. IEEE Computer Graphics and Applications 42, 116–122. doi:10.1109/MCG.2021.3132004.

Dai, R., Brell-Cokcan, S. (2022). Digital twins as education support in construction: A first development framework based on the reference construction site aachen west. Construction Robotics 6, 75–83. doi:10.1007/s41693-022-00070-7.

Driessen, J.E.P., Vos, D.A.C., Smeets, I., Albers, C.J. (2022). Misleading graphs in context: Less misleading than expected. PLOS ONE 17, e0265823. doi:10.1371/journal.pone.0265823.

Gold, J.M. (2014). Information processing correlates of a size-contrast illusion. Frontiers in Psychology 5, 142. doi:10.3389/fpsyg.2014.00142.

Guess, A.M., Lerner, M., Lyons, B., Montgomery, J.M., Nyhan, B., Reifler, J., Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the united states and india. Proceedings of the National Academy of Sciences 117, 15536–15545. doi:10.1073/pnas.1920498117.

Jin, M.X., Rajan, S., Bicas, C.E.G., Hao, M., Dong, L., Mufson, B., Hafiz, I. (2021). Novel validated index for the measurement of disinformation susceptibility at the county level. Cureus 13, e15305. doi:10.7759/cureus.15305.

Kabashkin, I. (2025). Ai-based digital twins of students: A new paradigm for competency-oriented learning transformation. Information 16, 846. doi:10.3390/info16100846.

Katsiroumpa, A., Moisoglou, I., Mangoulia, P., Konstantakopoulou, O., Gallos, P., Tsiachri, M., Galanis, P. (2025). The online misinformation susceptibility scale: Development and initial validation. Healthcare 13, 2252. doi:10.3390/healthcare13172252.

Lauer, C., O'Brien, S. (2020). The deceptive potential of common design tactics used in data visualizations, in: Proceedings of the 38th ACM International Conference on Design of Communication, ACM. doi:10.1145/3380851.3416762.

Lee, H., Yang, H. (2023). Digital twinning and optimization of manufacturing process flows. Journal of Manufacturing Science and Engineering 145, 111008. doi:10.1115/1.4063234.

Lisnic, M., Polychronis, C., Lex, A., Kogan, M. (2023). Misleading beyond visual tricks: How people actually lie with charts, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, ACM. doi:10.1145/3544548.3580910.

Loftus, E.F., Palmer, J.C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. Journal of Verbal Learning and Verbal Behavior 13, 585–589. doi:10.1016/S0022-5371(74)80011-3.

Lu, J., Yang, Z., Zheng, X., Wang, J., Kiritsis, D. (2022). Exploring the concept of cognitive digital twin from model-based systems engineering perspective. The International Journal of Advanced Manufacturing Technology 121, 5835–5854. doi:10.1007/s00170-022-09610-5.

Maertens, R., Götz, F.M., Golino, H.F., Roozenbeek, J., van der Linden, S. (2024). The misinformation susceptibility test (MIST): A psychometrically validated measure of news veracity discernment. Behavior Research Methods 56, 1863–1899. doi:10.3758/s13428-023-02124-2.

Mruczek, R.E.B., Blair, C.D., Strother, L., Caplovitz, G.P. (2015). The dynamic Ebbinghaus: Motion dynamics greatly enhance the classic contextual size illusion. Frontiers in Human Neuroscience 9, 77. doi:10.3389/fnhum.2015.00077.

Ratchita, B., Anandhi, S. (2026). Multimodal human digital twin AI for real time productivity and fatigue intelligence. International Journal of Science, Strategic Management and Technology 02, 1–9. doi:10.55041/ijsmt.v2i3.206.

Rho, J., Rau, M.A. (2025). Exploring educational approaches to addressing misleading visualizations. Educational Psychology Review 37, 14. doi:10.1007/s10648-025-09988-0.

Rho, J., Rau, M.A., Zhu, X. (2026). Comparing the deceptive impact of misleading data visualizations: Implications for adaptive data literacy support in computer-based learning environments. Computers & Education 242, 105508. doi:10.1016/j.compedu.2025.105508.

Shapiro, A.G., Todorovic, D. (2017). The Oxford Compendium of Visual Illusions. Oxford University Press. doi:10.1093/acprof:oso/9780199794607.001.0001.

Sturt, J., Ali, S., Robertson, W., Metcalfe, D., Grove, A., Bourne, C., Bridle, C. (2012). Neurolinguistic programming: A systematic review of the effects on health outcomes. British Journal of General Practice 62, e757–e764. doi:10.3399/bjgp12X658287.

Tao, F., Qi, Q., Wang, L., Nee, A.Y.C. (2019). Digital twins and cyber–physical systems toward smart manufacturing and industry 4.0: Correlation and comparison. Engineering 5, 653–661. doi:10.1016/j.eng.2019.01.014.

Todorović, D., Jovanović, L. (2018). Is the Ebbinghaus illusion a size contrast illusion? Acta Psychologica 185, 180–187. doi:10.1016/j.actpsy.2018.02.011.

Tversky, A., Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science 211, 453–458. doi:10.1126/science.7455683.

Xu, W., Yang, H., Ji, Z., Ba, M. (2024). Cognitive digital twin-enabled multi-robot collaborative manufacturing: Framework and approaches. Computers & Industrial Engineering 194, 110418. doi:10.1016/j.cie.2024.110418.

Zalizko, V. (2025). Spline-guided segmentation of handwritten physico-mathematical documents for improved OCR accuracy. Artificial Intelligence in Education: Ukraine and the World doi:10.21203/rs.3.rs-8471729/v1. preprint.

Zhou, C., Wang, Q., Yu, M., Yue, X., Lu, R., Li, J., Zhou, Y., Zhang, S., Zhou, J., Lam, W. (2025). The essence of contextual understanding in theory of mind: A study on question answering with story characters, in: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics. pp. 22612–22631. doi:10.18653/v1/2025.acl-long.1103.