International Society of the Learning Sciences

Playing with Intelligence: Evolving Epistemic Criteria in a Turing Test–Like Game

Zhongyu Wang, Florida State University, zw23@fsu.edu

Xiang Li, Lingfei Cao, Zhichun Liu
xianglix@connect.hku.hk, lingfei.cao@connect.hku.hk, lukasliu@hku.hk
The University of Hong Kong

Abstract: This study investigates how learners developed epistemic understanding of artificial intelligence through participation in a multiplayer Turing test-like game. Informed by the framework of epistemic cognition, we analyzed how 32 middle school students' game strategies and epistemic criteria evolved across four gameplay sessions. Findings show that students' epistemic criteria for identifying AI shifted from focusing on surface linguistic cues to socially and contextually grounded judgments. The study illustrates how game-based environments can foster the development of epistemic cognition through reflective reasoning.

Introduction and theoretical framework

As generative artificial intelligence (GenAI) becomes increasingly capable of mimicking human generations, figuring out the ambiguity involved in students' conceptualization of AI has become a necessary part of AI literacy (Long & Magerko, 2020). While most educational efforts focus on technical skills or conceptual knowledge, such as understanding algorithms, data processing, or ethical implications (Li et al., 2023; Ng et al., 2021; Zawacki-Richter et al., 2019), students' deeper epistemic understanding, that is, what AI means to them, was often overlooked in existing efforts.

Existing research has shown that game-based environments are promising for engaging students with epistemic understanding in authentic, problem-solving contexts (Barzilai, 2017). In this study, we designed a multiplayer Turing test-like game, *Occultation*, with the aim of promoting students' epistemic understanding (Chinn et al., 2011) about AI through investigating the boundaries between human and AI.

In addition, this study is grounded in the framework of epistemic cognition, which examines how individuals think about knowledge, knowing, and justification (Chinn et al., 2011). In this context, students' observable game strategies to distinguish human and AI function as reliable processes that reveal their underlying epistemic criteria (evaluative standards) (Barzilai & Chinn, 2025) for identifying intelligence. Accordingly, this study addresses two research questions:

(1) How did students' use of game strategies reveal their epistemic criteria for conceptualizing AI?
(2) How did students' game strategies and epistemic criteria evolve across four gameplay sessions?

Method

The study took place during a six-day summer program at a middle school, with 32 seventh-graders (15 males and 17 females). The program consisted of six one-hour sessions (see Figure 1); the first and final sessions were dedicated to pre- and post-tests, while sessions 2 through 5 were reserved for gameplay. The game featured two modes: Interrogation Mode, where groups of human players identified a hidden AI, and Traitor Mode, where specific participants used a private prompt-engineering interface to help an AI bot survive the interrogation. Across the sessions, the number of traitors increased to introduce greater ambiguity and social complexity. Data analysis adopted a theory-informed interpretive coding approach (Braun & Clarke, 2022) and involved coding 21 group chat histories to identify 21 distinct strategies, which were then axial-coded into seven higher-order epistemic criteria.

Findings

RQ1. Epistemic Criteria for Identifying AI

Analysis of students' chat histories revealed three overarching types of epistemic criteria for identifying AI: Knowledge, Ability, and Identity. Knowledge-related criteria focused on conceptualizing intelligence based on what knowledge it possesses. This included Personal Knowledge, where students shared daily experiences or checked social relationships to identify the hidden AI. For instance, asking "Guys, has anything good

 © ISLS

Figure 1
Data Collection Procedures

Date	Day 01	Day 02	Day 03	Day 04	Day 05	Day 06
Activity	Pre-test	Interrogation Mode	Traitor mode + Interrogation Mode	Traitor mode + Interrogation Mode	Traitor mode + Interrogation Mode	Post-test
Total number of interrogators	N/A	26	26	16	6	N/A
Total number of traitors	N/A	6	6	16	26	N/A

happened?" served as a strategy to distinguish agents based on their possession of real-life experiences. Students also relied on Knowledge About Community Culture and Norms, assuming AI would be unfamiliar with local humor, internet slang, or the implicit completion of song lyrics common in their social circles. Finally, Subject Knowledge was used as a reverse-indicator; students believed AI could answer complex domain-specific questions, such as French or advanced math, which their human peers likely could not.

Ability-related criteria concerned the communicative, contextual, and embodied performances that the intelligence enacted. Linguistic styles focused on form-level features; students analyzed typos, response length, and speed to infer humanness, often concluding that "Al doesn't make typos" or that excessively long messages must be machine-generated. Contextual criteria involved testing whether an agent could follow conversation flow or react appropriately to "nonsense" behaviors, such as spam messages or swear words used as memes. Physical actions reflected a belief that AI lacks embodiment; students employed metagaming strategies like asking an agent to "raise your hand" to confirm a physical presence in the room.

Identity-related criteria encompassed self-awareness and personal distinctiveness as qualities unique to humans. Under Self-identification, students used direct questioning like "Are you AI?" to observe responses to topics requiring self-consciousness. Gender and Personal Identity were also significant; students viewed gender as a distinctly human attribute and used the disclosure of names and social roles, such as "I'm Alex from Class 1", to express situated selves that they believed AI could not possess.

RQ2. Evolution of strategies and reasoning

Across the four gameplay sessions, students' strategies evolved from broad experimentation to evidence-based refinement. In early sessions, exploratory actions reflected curiosity and uncertainty about what generated meaningful evidence of humanness. By the third session, students began to abandon ineffective strategies that AI could mimic, such as monitoring emojis or punctuation, and focused on context-specific questioning. Conceptually, their epistemic criteria shifted from reliance on surface-level cues, such as response speed or typos, to contextually and socially grounded reasoning. By the final session, intelligence was defined less by linguistic precision and more by relational authenticity and the ability to demonstrate a shared understanding within their specific social group.

Discussion and conclusion

This study revealed how students' epistemic criteria for conceptualizing AI developed through active interaction with both human and AI agents. The findings show that, after rounds of testing, students refined not only how they tested for intelligence but also what they considered as evidence of it, progressing from behavioral observations to socially grounded interpretations. In summary, the study explores and describes a lived experience of epistemic growth, illustrating how students learn to reason with and about AI.

Theoretically, this study specifies the concept of epistemic criteria in the context of conceptualizing intelligence rather than knowledge in general. From the perspective of game design, it shows that game-based uncertainty can trigger epistemic development. Constant improvement in AI across four sessions introduced new ambiguity in identifying AI, prompting students to reflect on their reasoning, push at the boundaries of existing beliefs, and transform gameplay into a process of epistemic negotiation. This implies that epistemic changes occurred through iterative sensemaking. Methodologically, the study demonstrates how games can capture epistemic change in action rather than through retrospective reports. In-game actions reveal how students generate, test, and revise ideas in real time, offering dynamic data for studying epistemic reasoning.

 © ISLS

References

Barzilai, S. (2017). "Half-reliable": A qualitative analysis of epistemic thinking in and about a digital game. *Contemporary Educational Psychology*, *51*, 51–66. https://doi.org/10.1016/j.cedpsych.2017.06.004

Barzilai, S., & Chinn, C. A. (2018). On the goals of epistemic education: Promoting apt epistemic performance. *Journal of the Learning Sciences, 27*(3), 353–389. https://doi.org/10.1080/10508406.2017.1392968

Barzilai, S., & Chinn, C. A. (2025). How do source evaluation criteria develop? A microgenetic study of growth of epistemic ideals. *Computers in Human Behavior, 172*, 108729. https://doi.org/10.1016/j.chb.2025.108729

Braun, V., & Clarke, V. (2022). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing researcher. *International Journal of Transgender Health, 24*(1), 1–6. https://doi.org/10.1080/26895269.2022.2129597

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist, 46*(3), 141–167. https://doi.org/10.1080/00461520.2011.587722

Li, L., Yu, F., & Zhang, E. (2024). A systematic review of learning task design for K-12 AI education: Trends, challenges, and opportunities. *Computers and Education: Artificial Intelligence, 6,* 100217. https://doi.org/10.1016/j.caeai.2024.100217

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, USA,* (pp. 1–16). https://doi.org/10.1145/3313831.3376727

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2,* 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, *55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International Journal of Educational Technology in Higher Education, 16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0