

THIẾT KẾ VÀ THỬ NGHIỆM BÀI KIỂM TRA KIẾN THỨC CƠ BẢN VỀ TRÍ TUỆ NHÂN TẠO CHO HỌC SINH TRUNG HỌC CƠ SỞ

DESIGN AND PILOT TESTING OF A BASIC ARTIFICIAL INTELLIGENCE KNOWLEDGE ASSESSMENT FOR
LOWER SECONDARY STUDENTS

Lê Quang Quân^{1,2,+},
Lê Thái Hưng²,
Lê Anh Vinh¹

¹Viện Khoa học Giáo dục Việt Nam;

²Trường Đại học Giáo dục - Đại học Quốc gia Hà Nội

+Tác giả liên hệ • Email: quanlq@vnies.edu.vn

Article history

Received: 04/02/2026

Accepted: 26/02/2026

Published: 20/4/2026

Keywords

Artificial Intelligence (AI),
AI education, lower
secondary students,
assessment, test,
Response Theory (IRT), pilot
study

ABSTRACT

The rapid development of artificial intelligence (AI) has created a growing need to implement AI education for general education. In Vietnam, recent educational policies have emphasized the integration of AI education into schools. This study designed and piloted an assessment instrument to measure basic AI knowledge among lower secondary students. The proposed instrument consists of an 18-item test developed based on a pilot AI education content framework for general education students. Data collected from 468 students were analyzed using the Rasch model within the framework of item response theory (IRT). The results indicate good model fit and high internal reliability (Cronbach's Alpha = 0.91). The findings confirm the feasibility and value of the instrument for diagnosing lower secondary students' initial understanding of AI, while also supporting the implementation of AI education in schools. The study further suggests directions for future research toward a more comprehensive assessment of AI competencies among general education students in Vietnam.

1. Mở đầu

Trong những năm gần đây, sự phát triển nhanh chóng của trí tuệ nhân tạo (AI) đã khẳng định tiềm năng ứng dụng rộng rãi, khiến cho công nghệ này dần trở nên phổ biến và thành một phần quan trọng của nhiều lĩnh vực trong đời sống KT-XH. UNESCO (2025) nhận định AI đang làm thay đổi cách chúng ta học tập, giảng dạy và hiểu về thế giới xung quanh. Với lĩnh vực giáo dục, các nghiên cứu gần đây không chỉ dừng lại ở việc khám phá khả năng ứng dụng AI mà còn quan tâm nhiều đến khía cạnh phát triển năng lực AI cho người học, đặc biệt là ở bậc phổ thông - nơi hình thành nền tảng nhận thức và định hướng nghề nghiệp tương lai. Theo UNESCO (2022), ít nhất 11 quốc gia đã chính thức phê duyệt chương trình giảng dạy AI và 04 quốc gia khác đang trong giai đoạn hoàn thiện. Đến nay, nghiên cứu tổng quan cho thấy nhiều quốc gia đã và đang triển khai nội dung giáo dục AI cho HS phổ thông trên diện rộng, tiêu biểu như Trung Quốc, Hàn Quốc, Nhật Bản, Á-rập Xê-út, Úc, Hoa Kỳ (Viện Khoa học Giáo dục Việt Nam, 2025).

Theo Lê Anh Vinh và Trần Mỹ Ngọc (2024), AI là công cụ duy nhất có khả năng sẽ nhanh chóng rút ngắn khoảng cách về ưu thế cạnh tranh giữa giáo dục Việt Nam và các nước đang phát triển. Ở Việt Nam, vấn đề đưa nội dung giáo dục AI vào nhà trường đã được cụ thể hoá trong nhiều chủ trương, chính sách và văn bản chỉ đạo, đảm bảo tính thống nhất, đồng bộ và có hệ thống. Quyết định số 127/QĐ-TTg ngày 26/01/2021 của Thủ tướng Chính phủ ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng AI đến năm 2030 xác định mục tiêu xây dựng chương trình đào tạo AI phù hợp từng cấp học, đưa kiến thức AI vào Chương trình giáo dục phổ thông (GDPT), chương trình đào tạo GV (Thủ tướng Chính phủ, 2021); Nghị quyết số 71-NQ/TW ngày 22/8/2025 về đột phá phát triển GD-ĐT đặt ra nhiệm vụ "nghiên cứu đưa kiến thức về năng lực số, AI vào Chương trình GDPT" (Ban Chấp hành Trung ương, 2025). Gần đây, Bộ GD-ĐT (2025) đã ban hành quyết định số 3439/QĐ-BGDĐT ngày 15/12/2025 về Khung nội dung thí điểm giáo dục AI cho HS phổ thông và triển khai các hoạt động tập huấn, thí điểm.

Để triển khai giáo dục AI một cách hiệu quả, việc đánh giá mức độ sẵn sàng và hiểu biết ban đầu của HS là yêu cầu thiết yếu. Trong bối cảnh đó, việc phát triển một công cụ đánh giá hiểu biết căn bản về AI của HS các cấp đóng vai trò then chốt: (1) Giúp nhận diện mức độ hiểu biết hiện tại; (2) Hỗ trợ đánh giá sự sẵn sàng của người học trước khi tiếp nhận nội dung giáo dục AI; (3) Cung cấp dữ liệu thực nghiệm phục vụ việc thiết kế chương trình, học liệu và phương pháp giảng dạy phù hợp. Trên cơ sở đó, bài báo giới thiệu quá trình thiết kế, thử nghiệm một bài kiểm tra

đánh giá hiểu biết căn bản về AI của HS THCS cùng các phân tích theo lí thuyết ứng đáp câu hỏi (Item Response Theory - IRT). Công cụ được xây dựng dựa trên định hướng của Khung nội dung thí điểm giáo dục AI cho HS phổ thông, tập trung vào các nội dung cơ bản về kĩ thuật và ứng dụng AI ở cấp THCS. Kết quả nghiên cứu cung cấp bằng chứng thực nghiệm ban đầu về tính khả thi của công cụ đối với mục tiêu khảo sát và triển khai giáo dục AI ở bậc phổ thông, đồng thời đề xuất các khuyến nghị cho việc phát triển, mở rộng và hoàn thiện trong các giai đoạn tiếp theo.

2. Kết quả nghiên cứu

2.1. Cơ sở lí thuyết

2.1.1. Cơ sở lí thuyết về phạm vi nghiên cứu và nội dung đánh giá

Những năm gần đây, thuật ngữ “năng lực hiểu biết về AI” (AI literacy) được nhắc đến như một khái niệm phổ biến trong các nghiên cứu về AI trong giáo dục, với nhiều cách tiếp cận theo mục tiêu và bối cảnh triển khai khác nhau. Mills và cộng sự (2024) đã tổng hợp và phân tích các khung lí thuyết dựa trên nhiều nghiên cứu khác nhau và thực tiễn giáo dục, từ đó đề xuất cách tiếp cận năng lực hiểu biết về AI bao gồm các kiến thức và kĩ năng cho phép con người hiểu một cách có phân biệt, đánh giá và sử dụng các hệ thống AI một cách an toàn và hiệu quả, nhằm tham gia có trách nhiệm trong một thế giới ngày càng số hoá. Kết quả nghiên cứu nhấn mạnh ba phương thức tham gia có mối liên hệ chặt chẽ với AI là hiểu (understand), đánh giá (evaluate) và sử dụng (use), được xem là nền tảng cho việc hình thành năng lực hiểu biết về AI trong GDPT.

Theo Lê Anh Vinh và cộng sự (2025), quá trình nghiên cứu xây dựng khung nội dung giáo dục AI ở Việt Nam tiếp cận theo phạm vi rộng hơn, ở mức độ tổng thể là giáo dục phát triển năng lực AI (AI competency), từ đó thích ứng với bối cảnh giáo dục trong nước và đưa ra cấu trúc theo bốn thành phần năng lực, với yêu cầu cần đạt về năng lực AI đặc thù đối với HS cấp THCS thể hiện ở bảng 1 dưới đây.

Bảng 1. Yêu cầu cần đạt về năng lực AI đặc thù đối với HS cấp THCS (Nguồn: Bộ GD-ĐT, 2025)

Thành phần năng lực	Biểu hiện
NLa: Tư duy lấy con người làm trung tâm	Hiểu vai trò của con người trong thiết kế, vận hành, sử dụng AI; Biết rằng con người chịu trách nhiệm với các phản hồi, tác động AI; Nhận diện được, phân tích tình huống sử dụng AI cho mục đích đúng đắn; Biết đề xuất cách kết hợp AI với yếu tố con người giải quyết vấn đề xã hội
NLb: Đạo đức AI	Nêu được các nguyên tắc đạo đức cơ bản của AI: không gây hại, không thiên kiến, công bằng, minh bạch; Biết áp dụng các nguyên tắc đó khi học tập, làm việc với công cụ AI; Biết đánh giá sản phẩm AI có nguy cơ xâm phạm quyền con người, đề xuất giải pháp cải thiện
NLc: Các kĩ thuật và ứng dụng AI	Hiểu được khái niệm dữ liệu, thuật toán, mô hình AI; Biết dùng một số công cụ AI đơn giản phục vụ học tập, dự án nhỏ; Biết kết hợp nhiều công cụ AI để tạo sản phẩm số có ý nghĩa (video, thuyết trình, mô phỏng)
NLd: Thiết kế hệ thống AI	Xác định được các tình huống thực tiễn có thể và nên ứng dụng AI; Tham gia vào quá trình lập kế hoạch thiết kế hệ thống AI đơn giản thông qua việc xác định mục tiêu, lựa chọn dữ liệu và mô phỏng hoạt động với công cụ có sẵn; Đánh giá được kết quả và đề xuất cách cải thiện ở mức độ cơ bản

Trong nghiên cứu này, nhóm tác giả giới hạn phạm vi tiếp cận ở mức độ hẹp hơn, tập trung vào khía cạnh kiến thức cơ bản về AI của HS THCS, như một bước khởi đầu cần thiết trước khi xem xét đánh giá đầy đủ và toàn diện các khía cạnh năng lực AI ở mức độ cao hơn. Theo đó, công cụ xác định hướng đến đo lường mức độ HS có thể nhận diện được các ứng dụng phổ biến của AI trong học tập và đời sống, hiểu được một số khái niệm, nguyên lí ở mức căn bản và nhận thức được cách thức AI được sử dụng trong các tình huống quen thuộc, phù hợp với đặc điểm nhận thức của HS THCS. Cách tiếp cận này nhấn mạnh các thành tố về nhận thức và hiểu biết ban đầu mà chưa bao hàm đầy đủ các yêu cầu về kĩ năng thực hành phức hợp hay khả năng phân biệt sâu theo các chuẩn mực đạo đức - xã hội ở mức độ toàn diện.

2.1.2. Cơ sở lí thuyết về thiết kế công cụ và đo lường

Theo Zhou và cộng sự (2025), các nghiên cứu về đánh giá năng lực/hiểu biết về AI của HS hiện nay có nhiều cách thức triển khai khác nhau, trong đó có 05 phương pháp tiếp cận chính: (1) Sử dụng bài kiểm tra kiến thức - hiểu biết về AI; (2) Khảo sát tự đánh giá; (3) Đánh giá theo các tiêu chí (rubric) thông qua dự án, sản phẩm của HS; (4) Quan sát lớp học; (5) Phỏng vấn. Nghiên cứu của Wang và cộng sự (2023) chỉ ra rằng năng lực hiểu biết AI của HS THCS có thể được đánh giá dựa vào các câu hỏi đóng, theo hình thức câu hỏi trắc nghiệm, câu hỏi đa lựa chọn

và câu hỏi thang đo. Do đó, trong nghiên cứu này, nhóm tác giả lựa chọn hình thức bài kiểm tra gồm các câu hỏi trắc nghiệm và đa lựa chọn để đo lường kiến thức cơ bản về AI của HS THCS.

Về mặt đo lường, nghiên cứu tiếp cận theo lý thuyết ứng đáp câu hỏi (Item Response Theory - IRT) như một cơ sở lý thuyết nền tảng cho việc thiết kế và phân tích công cụ đánh giá. IRT cho phép mô tả mối quan hệ giữa năng lực tiềm ẩn của người học và xác suất trả lời đúng từng câu hỏi, đồng thời cung cấp các chỉ số quan trọng về độ khó, độ phân biệt và mức độ phù hợp của từng mục hỏi, độc lập tương đối với mẫu khảo sát (Baker, 2001). Việc vận dụng IRT với mô hình Rasch trong nghiên cứu này nhằm đảm bảo tính nhất quán đo lường và cung cấp bằng chứng thực nghiệm cho chất lượng của bài kiểm tra kiến thức cơ bản về AI.

2.2. Phương pháp nghiên cứu

Để đảm bảo tính toàn diện và khoa học trong việc phát triển, đánh giá chất lượng bài kiểm tra, nghiên cứu được thiết kế theo cách tiếp cận hỗn hợp với sự kết hợp chặt chẽ giữa các phương pháp nghiên cứu định tính và định lượng. Ở giai đoạn thiết kế công cụ, phương pháp định tính được sử dụng làm chủ đạo, bao gồm các phương pháp nghiên cứu tổng quan, phân tích hệ thống cơ sở lý luận để đề xuất cách tiếp cận khung năng lực, khung đánh giá; phương pháp thảo luận nhóm tập trung nhằm xác lập ma trận bài kiểm tra; và phương pháp chuyên gia để thẩm định độ giá trị nội dung của từng mục hỏi. Tiếp sau đó, trong giai đoạn thử nghiệm và đánh giá thực nghiệm, nghiên cứu vận dụng phương pháp định lượng thông qua việc triển khai bài kiểm tra và sử dụng lý thuyết ứng đáp câu hỏi với mô hình Rasch để phân tích các thuộc tính của bộ công cụ. Quy trình nghiên cứu được cụ thể hoá qua 04 bước như sau:

Bước 1. Nghiên cứu tài liệu và xác lập ma trận nội dung thông qua thảo luận nhóm tập trung. Nhóm tác giả đã tiến hành phân tích tổng quan các bài công bố quốc tế về đánh giá năng lực/hiểu biết về AI và nghiên cứu, đối chiếu với Khung nội dung thí điểm giáo dục AI do Bộ GD-ĐT ban hành để hình thành dự thảo nội dung. Sau đó, một buổi thảo luận nhóm tập trung được thực hiện với sự tham gia của 05 thành viên gồm các chuyên gia về công nghệ giáo dục, công nghệ thông tin, đo lường và đánh giá trong giáo dục cùng GV môn Tin học cấp THCS. Hoạt động này nhằm thống nhất giới hạn nội dung, xác lập cấu trúc ma trận bài kiểm tra và mô tả diễn giải kèm ví dụ minh hoạ để xây dựng bộ câu hỏi chi tiết.

Bước 2. Thiết kế chi tiết mục hỏi và thẩm định qua phương pháp chuyên gia. Dựa trên ma trận đã được thống nhất, nhóm nghiên cứu tiến hành biên soạn chi tiết 18 câu hỏi cho bài kiểm tra. Để hoàn thiện công cụ, nhóm sử dụng phương pháp chuyên gia để thu thập ý kiến đóng góp độc lập từ 05 nhà khoa học và GV cốt cán, thuộc các lĩnh vực: công nghệ giáo dục, công nghệ thông tin, ngôn ngữ học, tâm lý học và GV môn Tin học THCS. Các chuyên gia đánh giá mức độ phù hợp của bộ câu hỏi dựa trên các tiêu chí về ma trận đã thống nhất, tính khoa học, đặc điểm nhận thức của HS THCS và sự tường minh của ngôn ngữ. Quy trình này đảm bảo bộ công cụ đạt được sự đồng thuận cao về chất lượng nội dung và hình thức trước khi đưa vào thử nghiệm thực tế.

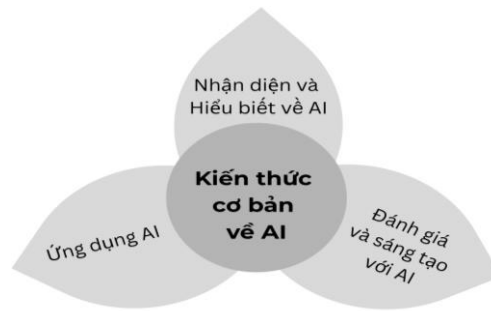
Bước 3. Triển khai khảo sát thực nghiệm và thu thập dữ liệu. Hoạt động thực nghiệm được thực hiện thông qua bài kiểm tra trực tuyến thiết kế trên nền tảng Google Forms, triển khai vào giữa tháng 12/2025. Để đảm bảo tính đại diện và đa dạng của mẫu, nghiên cứu được triển khai tại 03 trường THCS thuộc địa bàn TP Hà Nội và tỉnh Nghệ An. Dữ liệu sau khi thu thập được tiến hành làm sạch, loại bỏ các phiếu trả lời tương ứng với các HS xác nhận chưa từng biết đến khái niệm/thuật ngữ về AI. Tập dữ liệu chính thức gồm 468 HS được mã hoá và chuẩn bị cho phân tích định lượng chuyên sâu. Mô tả thống kê hậu nghiệm cho thấy mẫu có sự phân bố tương đối đồng đều theo khối lớp và giới tính.

Bước 4. Phân tích các thuộc tính của bộ công cụ theo mô hình Rasch. Dữ liệu thực nghiệm được xử lý bằng phần mềm Conquest dựa trên mô hình của lý thuyết ứng đáp câu hỏi. Nhóm nghiên cứu tập trung phân tích các chỉ số phù hợp, giá trị độ khó, độ tin cậy, độ phân biệt theo các hướng dẫn của Baker (2001), Bond và cộng sự (2021). Kết quả phân tích ở bước này cung cấp minh chứng khoa học về tính khả thi, độ tin cậy và giá trị cấu trúc của bài kiểm tra kiến thức cơ bản về AI đã được thiết kế.

2.3. Kết quả phân tích và thảo luận

2.3.1. Ma trận bài kiểm tra

Trên cơ sở tổng quan tài liệu và nghiên cứu khung nội dung thí điểm giáo dục AI cho HS phổ thông, với định hướng khu trú vào kiến thức cơ bản về AI, cấu trúc nội dung của bài kiểm tra đã được xác lập thông qua phương pháp thảo luận nhóm tập trung với sự tham gia của các chuyên gia giáo dục, chuyên gia công nghệ và GV môn Tin học. Kết quả thảo luận đã đi đến thống nhất rằng việc đánh giá hiểu biết ban đầu về AI ở cấp THCS cần phản ánh tiến trình nhận thức của người học theo một lộ trình từ nhận diện đến vận dụng thực tế. Theo đó, nội dung kiến thức cơ bản về AI trong nghiên cứu này được cấu trúc thành ba khía cạnh cốt lõi, đóng vai trò là khung định hướng để xây dựng ma trận bài kiểm tra chi tiết (hình 1).



Hình 1. Minh họa cấu trúc các khía cạnh nội dung kiến thức cơ bản về AI đối với HS THCS (Nguồn: Tác giả)

Ma trận bài kiểm tra được thiết kế nhằm đảm bảo sự phân bố khoa học và hợp lý, gồm 18 câu hỏi theo từng khía cạnh nội dung tương ứng với các mức độ nhận thức. Cụ thể, 03 khía cạnh nội dung chính bao gồm: (1) Nhận diện và Hiểu biết về AI, tập trung đo lường các kiến thức nền tảng và khả năng nhận biết các ứng dụng phổ biến của AI phù hợp với trải nghiệm của HS THCS; (2) Ứng dụng AI, phản ánh mức độ hiểu biết của HS về cách thức sử dụng công nghệ AI trong các bối cảnh thực tiễn quen thuộc; (3) Đánh giá và sáng tạo với AI, được tiếp cận ở mức độ nhận thức ban đầu, nhấn mạnh vào khả năng vận dụng kiến thức để đưa ra các phán đoán, nhận định hoặc đề xuất các ý tưởng giải quyết vấn đề. Sự phân bố chi tiết các mục hỏi theo ma trận được trình bày như trong bảng dưới đây.

Bảng 2. Ma trận bài kiểm tra kiến thức căn bản về AI của HS THCS

	Nhận biết	Thông hiểu	Vận dụng	Tổng số câu
Nhận diện và Hiểu biết về AI	5	4		9
Ứng dụng AI		3	2	5
Đánh giá và sáng tạo với AI		1	3	4
Tổng số câu	5	8	5	18

2.3.2. Phân tích dữ liệu thử nghiệm theo lý thuyết ứng đáp câu hỏi

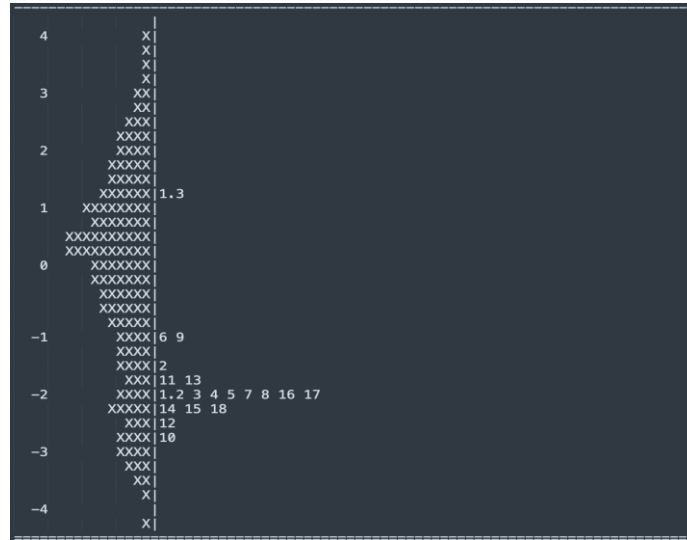
- *Mức độ phù hợp với mô hình và độ tin cậy của bài kiểm tra:* Theo Bond và cộng sự (2021), độ phù hợp của các mục (item fit) có thể được đánh giá thông qua chỉ số bình phương trung bình có trọng số (Weighted Mean Square - Infit MNSQ) với khoảng chấp nhận từ 0,7 đến 1,3. Kết quả phân tích cho thấy 17/18 câu hỏi có chỉ số MNSQ nằm trong khoảng này, thể hiện mức độ phù hợp tốt với mô hình Rasch. Riêng đối với câu hỏi 1, chỉ số Infit MNSQ = 1,50 nằm ngoài ngưỡng tối ưu nhưng vẫn có thể được xếp vào ngưỡng chấp nhận được (acceptable range), việc sai lệch ở chỉ số này không đồng nghĩa với việc mục đo phải bị loại bỏ, mà cần được xem xét trong mối liên hệ với tính phù hợp về nội dung và mục đích đo lường của công cụ (Bond và cộng sự, 2021). Ở câu hỏi này, thí sinh được chấm theo ba mức điểm: 1 điểm nếu trả lời đúng 1-3 ý, 2 điểm nếu trả lời đúng 4-6 ý và 3 điểm nếu trả lời đúng 7-8 ý. Việc giá trị MNSQ cao hơn so với các câu hỏi còn lại là tín hiệu cho thấy cần phân tích sâu hơn đối với dạng mục hỏi đa mức điểm này ở bước tiếp theo. Về độ tin cậy, hệ số Cronbach's Alpha của bài kiểm tra đạt 0,91 cho thấy tính nhất quán nội tại cao. Giá trị độ tin cậy tách biệt của câu hỏi (Item Separation Reliability) đạt 0,937, cho thấy bài kiểm tra có khả năng phân biệt tốt giữa các câu hỏi theo độ khó. Đồng thời, độ tin cậy tách biệt của HS (EAP/PV Person Reliability) đạt 0,826 cho thấy bài kiểm tra có khả năng phân biệt các nhóm HS theo năng lực một cách đáng tin cậy.

- *Độ khó và độ phân biệt của câu hỏi:* Được tổng hợp như trong bảng dưới đây:

Bảng 3. Tổng hợp giá trị độ khó và độ phân biệt của các câu hỏi

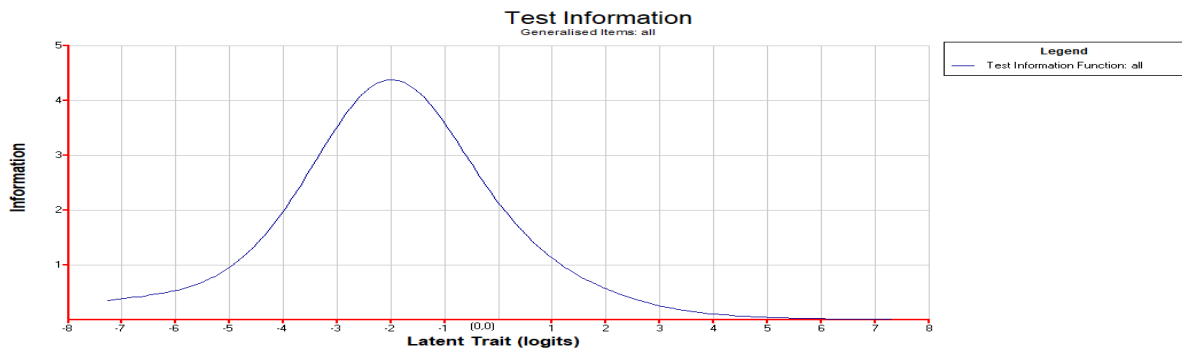
Câu	Độ khó	Độ phân biệt	Câu	Độ khó	Độ phân biệt
1.1	-7,03	0,52	9	-1,03	0,48
1.2	-2,10		10	-2,68	0,45
1.3	1,20		11	-1,75	0,62
2	-1,55	0,62	12	-2,54	0,72
3	-2,11	0,69	13	-1,88	0,55
4	-2,11	0,71	14	-2,28	0,73
5	-2,05	0,74	15	-2,32	0,69
6	-1,17	0,59	16	-1,96	0,66
7	-2,01	0,64	17	-1,98	0,61
8	-2,11	0,71	18	-2,20	0,72

Theo Baker (2001), các câu hỏi khi phân tích theo lý thuyết IRT nên có giá trị độ khó từ -3,0 đến 3,0 và có thể được chia thành 05 mức độ: Rất dễ (giá trị độ khó nhỏ hơn -2,0); Dễ (giá trị độ khó từ -2,0 đến dưới -0,5); Trung bình (từ -0,5 đến dưới 0,5); Khó (từ 0,5 đến dưới 2,0); Rất khó (từ 2,0 trở lên). Như vậy, ngoại trừ việc đạt điểm tối đa ở câu hỏi 1 được xếp vào mức khó thì các câu hỏi đều được đánh giá ở mức dễ và rất dễ. Điều này được nhận định là phù hợp với định hướng tiếp cận ban đầu của bài kiểm tra là đặt ra những câu hỏi ở mức hiểu biết căn bản về AI đối với HS THCS. Hình dưới đây mô tả một phần bảng phân bố độ khó của các câu hỏi và mức năng lực của thí sinh, với tỉ lệ $X \sim 3,3$ thí sinh.



Hình 2. Phân bố độ khó của các câu hỏi và mức năng lực của thí sinh

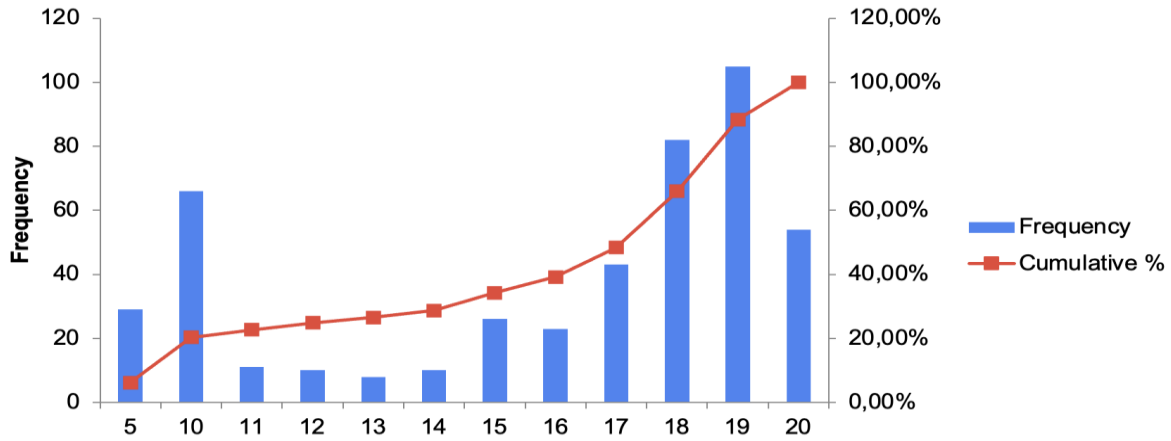
Mặc dù các câu hỏi dễ với yêu cầu ở mức hiểu biết căn bản nhưng chúng vẫn có khả năng phân biệt tốt. Các phương án nhiễu đều được đánh giá đảm bảo chất lượng với hệ số tương quan điểm - nhị phân (Point - Biserial correlation, Pt-Bis) âm, có ý nghĩa thống kê ($p < 0,001$). Kết quả phân tích cho thấy bài kiểm tra này phân biệt tốt nhất nhóm thí sinh có năng lực ở mức trung bình, điều này cũng được thể hiện rõ thông qua đường cong thông tin của bài kiểm tra (xem hình 3).



Hình 3. Đường cong thông tin cho cả đề thi

- *Thống kê mô tả kết quả bài làm của HS:* Kết quả làm bài của 468 HS cho thấy điểm trung bình đạt 15,35 điểm trên thang điểm tối đa 20, với độ lệch chuẩn 4,85, thể hiện mức độ phân tán điểm số tương đối lớn trong nhóm khảo sát. Phân bố điểm số nghiêng về phía các mức điểm cao: có khoảng 66% HS đạt từ 16 điểm trở lên, trong đó hơn 51% đạt mức từ 18 điểm trở lên. Đáng chú ý, có 54 HS đạt điểm tuyệt đối, chiếm tỉ lệ gần 12% tổng mẫu. Tuy nhiên, ở chiều ngược lại, gần 10% HS có kết quả dưới 10 điểm, phản ánh sự hiện diện rõ rệt của một nhóm HS có mức độ tiếp cận kiến thức còn hạn chế. Xét tổng thể, các kết quả mô tả cho thấy phần lớn HS đã thể hiện có nền tảng hiểu biết cơ bản về AI khá tốt. Tuy nhiên, sự phân hoá điểm số cho thấy tính không đồng đều trong năng lực giữa các HS. Việc tồn tại đồng thời các nhóm HS đạt điểm rất cao và rất thấp cho thấy sự khác biệt đáng kể về mức độ tiếp cận

kiến thức. Kết quả này mang hàm ý quan trọng đối với công tác triển khai giáo dục AI ở bậc THCS. Một mặt, nhóm đạt điểm cao cho thấy khả năng tiếp cận tốt và tiềm năng phát triển sâu hơn trong các nội dung về AI. Mặt khác, sự xuất hiện của nhóm HS có điểm số thấp hơn cho thấy nhu cầu hỗ trợ học tập bổ sung nhằm đảm bảo tính bao trùm và đồng bộ trong chương trình giáo dục AI.



Hình 4. Biểu đồ phân bố tần suất điểm số của HS

3. Kết luận

Kết quả nghiên cứu cho thấy bài kiểm tra kiến thức cơ bản về AI dành cho HS THCS được thiết kế và thử nghiệm đã bước đầu đáp ứng mục tiêu đo lường kiến thức nền tảng, thể hiện qua sự phù hợp của các câu hỏi với mô hình Rasch, các chỉ số độ phù hợp và độ tin cậy đạt ngưỡng chấp nhận, đồng thời có khả năng phân biệt mức độ hiểu biết giữa các nhóm HS. Giá trị của nghiên cứu thể hiện ở việc tích hợp một cách hệ thống các khung lý thuyết quốc tế với định hướng thực tiễn triển khai thí điểm giáo dục AI tại Việt Nam thông qua quy trình thiết kế đa bước, từ thảo luận nhóm chuyên gia đến thực nghiệm. Công cụ đề xuất có thể sử dụng trong các khảo sát diện rộng nhằm nhận diện mức độ hiểu biết cơ bản ban đầu về AI của HS, qua đó cung cấp dữ liệu tham chiếu cho nhà trường và cơ quan quản lý khi xây dựng kế hoạch triển khai giáo dục AI phù hợp. Trong bối cảnh giáo dục Việt Nam đang từng bước tiếp cận các định hướng và chính sách về giáo dục AI, nghiên cứu này cung cấp cơ sở thực nghiệm và đề xuất quy trình phát triển công cụ đo lường hiểu biết AI. Kết quả nghiên cứu góp phần bổ sung cơ sở khoa học cho việc triển khai đánh giá hiểu biết AI ở cấp THCS và có thể tham khảo cho các cấp học khác trong GDPT. Các nghiên cứu tiếp theo có thể tập trung phát triển bộ công cụ đánh giá năng lực AI ở mức độ cao hơn, mở rộng phạm vi nội dung đo lường và tích hợp các hình thức đánh giá dựa trên dự án, sản phẩm học tập. Đồng thời, việc vận dụng các lý thuyết và mô hình đo lường hiện đại sẽ giúp bảo đảm tính chuẩn hóa và khả năng so sánh giữa các nghiên cứu. Bên cạnh đó, cần tiếp tục phân tích các yếu tố ảnh hưởng đến mức độ tiếp cận và hiểu biết AI của HS, thử nghiệm các mô hình giáo dục AI phù hợp với điều kiện địa phương, nhằm thúc đẩy sự hình thành và phát triển năng lực AI cho HS phổ thông.

Lời cảm ơn: Bài báo được tài trợ bởi Chương trình học bổng đào tạo thạc sĩ, tiến sĩ trong nước của Quỹ Đổi mới sáng tạo Vingroup (VINIF), mã số VINIF.2024.ThS.18.

Tài liệu tham khảo

- Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation.
- Ban Chấp hành Trung ương (2025). *Nghị quyết số 71-NQ/TW ngày 22/8/2025 của Bộ Chính trị về đột phá phát triển giáo dục và đào tạo*.
- Bộ GD-ĐT (2025). *Quyết định số 3439/QĐ-BGDĐT ngày 15/12/2025 về việc Ban hành Khung nội dung thi điểm giáo dục trí tuệ nhân tạo cho học sinh phổ thông*.
- Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (Fourth edition). Routledge Taylor & Francis Group.

- Lê Anh Vinh, Đỗ Đức Lân, Nguyễn Hoài Nam, ..., Lương Đình Hải (2025). Khung năng lực Trí tuệ nhân tạo dành cho học sinh phổ thông Việt Nam. *Tạp chí Khoa học Giáo dục Việt Nam*, 21(7), 1-8.
- Lê Anh Vinh, Trần Mỹ Ngọc (2024). Tác động của Trí tuệ nhân tạo (AI) đối với hệ thống giáo dục toàn cầu và giáo dục Việt Nam. *Tạp chí Khoa học Giáo dục Việt Nam*, 20(5), 1-11.
- Mills, K., Ruiz, P., Lee, K., Coenraad, M., Fusco, J., Roschelle, J., & Weisgrau, J. (2024). *AI Literacy: A Framework to Understand, Evaluate, and Use Emerging Technology*. Digital Promise.
- Thủ tướng Chính phủ (2021). *Quyết định số 127/QĐ-TTg ngày 26/01/2021 ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng Trí tuệ nhân tạo đến năm 2030*.
- UNESCO. (2022). *K-12 AI curricula: A mapping of government-endorsed AI curricula*. UNESCO.
- UNESCO. (2025). *AI and the future of education: Disruptions, dilemmas and directions*. UNESCO. <https://doi.org/10.54675/KECK1261>
- Viện Khoa học Giáo dục Việt Nam (2025). *Báo cáo Kinh nghiệm quốc tế về xây dựng chương trình giáo dục phổ thông môn Trí tuệ nhân tạo và bài học cho Việt Nam*.
- Wang, X., Li, X., & Huang, J. (2023). Junior High School Artificial Intelligence Literacy: Connotation, Evaluation and Promotion Strategy. *Open Journal of Social Sciences*, 11(05), 33-49.
- Zhou, X., Li, Y., Chai, C. S., & Chiu, T. K. F. (2025). Defining, enhancing, and assessing artificial intelligence literacy and competency in K-12 education from a systematic review. *Interactive Learning Environments*, 33(10), 5766-5788. <https://doi.org/10.1080/10494820.2025.2487538>