THIẾT KẾ NHIỆM VỤ ĐÁNH GIÁ NĂNG LỰC LẬP LUẬN TOÁN HỌC CỦA HỌC SINH TRUNG HỌC PHỔ THÔNG TRONG BỐI CẢNH TRÍ TUỆ NHÂN TẠO TẠO SINH

Bùi Phương Linh
Trường THPT Nguyễn Trãi, xã Vũ Thư, tỉnh Hưng Yên
Email: linhphuong060602@gmail.com

Tóm tắt: *Nghiên cứu thiết kế nhiệm vụ đánh giá năng lực lập luận toán học của học sinh THPT trong bối cảnh AI tạo sinh. Theo khung Đánh giá dựa trên bằng chứng, lời giải AI được chuẩn hóa làm ngữ liệu để học sinh phát hiện sai sót, đưa ra căn cứ, xây dựng phản ví dụ và sửa chữa lập luận. Nghiên cứu sử dụng phương pháp thiết kế giáo dục, kết hợp định tính và định lượng với 125 học sinh. Kết quả cho thấy AI được sử dụng phổ biến, nhưng năng lực lập luận chưa đồng đều. Học sinh phát hiện sai sót tốt hơn xây dựng phản ví dụ và điều chỉnh lập luận. Nghiên cứu đề xuất chuẩn hóa ngữ liệu AI, thiết kế nhiệm vụ theo chuỗi hành động và sử dụng rubric bốn mức trong đánh giá năng lực*

Từ khóa: *năng lực lập luận toán học; AI tạo sinh; nhiệm vụ đánh giá; đánh giá dựa trên bằng chứng; học sinh trung học phổ thông.*

DESIGNING TASKS TO ASSESS HIGH SCHOOL STUDENTS' MATHEMATICAL REASONING ABILITY IN THE CONTEXT OF GENERATIVE ARTIFICIAL INTELLIGENCE

Abstract*: This study designed tasks to assess high school students' mathematical reasoning competence in the context of generative artificial intelligence. Based on the Evidence-Centered Design framework, standardized AI-generated solutions were used as assessment stimuli requiring students to identify errors, provide mathematical justification, construct counterexamples, and revise flawed reasoning. The study employed an educational design research approach, combining qualitative and quantitative methods with a sample of 125 students. The findings indicate that generative AI was widely used among students, although their mathematical reasoning competence varied considerably. Students performed better in identifying errors than in constructing counterexamples and revising arguments. The study recommends standardizing AI-generated materials, designing assessment tasks as a sequence of reasoning activities, and applying a four-level analytical rubric to improve the quality of competence-based assessment.*

Keywords: *mathematical reasoning competence; generative artificial intelligence; assessment tasks; Evidence-Centered Design; high school students.*

Nhận bài: 29/06/2026 Phản biện: 21/07/2026 Duyệt đăng: 26/07/2026

I. ĐẶT VẤN ĐỀ

Trong bối cảnh đổi mới giáo dục phổ thông theo định hướng phát triển phẩm chất và năng lực, kiểm tra, đánh giá môn Toán không chỉ xác định khả năng ghi nhớ kiến thức hoặc thực hiện phép tính mà còn phải phản ánh quá trình tư duy của học sinh. Chương trình Giáo dục phổ thông môn Toán năm 2018 xác định năng lực tư duy và lập luận toán học là một trong những năng lực cốt lõi. Năng lực này thể hiện ở khả năng phân tích dữ kiện, hình thành phán đoán, sử dụng căn cứ toán học, giải thích, chứng minh và điều chỉnh kết luận.

Sự phát triển nhanh chóng của trí tuệ nhân tạo tạo sinh như ChatGPT, Gemini và Copilot đang tạo ra những thay đổi đáng kể trong dạy học và đánh giá môn Toán. Các công cụ này có thể nhanh chóng tạo lời giải được trình bày mạch lạc nhưng vẫn có thể chứa sai sót về giả thiết, biến đổi hoặc quan hệ logic. Đồng thời, học sinh có thể giao toàn bộ nhiệm vụ cho AI mà không trực tiếp thực hiện hoạt động tư duy. OECD (2026) chỉ ra rằng việc hoàn thành nhiệm vụ bằng AI không đồng nghĩa với sự tiến bộ trong học tập nếu thiếu định hướng sư phạm phù hợp.

Do đó, cần thiết kế những nhiệm vụ yêu cầu học sinh phân tích, kiểm chứng và phản biện lời giải do AI tạo ra. Thông qua việc xác định lỗi, chỉ ra điều kiện còn thiếu, đưa ra phản ví dụ và sửa chữa lời giải, học sinh có cơ hội bộc lộ rõ hơn năng lực lập luận toán học.

Xuất phát từ yêu cầu trên, nghiên cứu được thực hiện tại Trường THPT Nguyễn Trãi, xã Vũ Thư, tỉnh Hưng Yên, nhằm khảo sát thực trạng sử dụng AI tạo sinh, xây dựng khung chỉ báo, thiết kế hệ thống nhiệm vụ và rubric phân tích bốn mức để đánh giá năng lực lập luận toán học; đồng thời thử nghiệm các nhiệm vụ với 125 học sinh thuộc các khối 10, 11 và 12. Kết quả nghiên cứu cung cấp cơ sở thực tiễn cho việc đổi mới kiểm tra, đánh giá môn Toán theo hướng chú trọng quá trình phân tích, kiểm chứng và điều chỉnh lập luận; qua đó

khuyến khích học sinh tư duy độc lập và sử dụng AI có trách nhiệm.

II. CƠ SỞ LÝ LUẬN

2.1. Năng lực lập luận toán học của học sinh THPT

Lập luận toán học là thành tố quan trọng của hoạt động học tập môn Toán, giúp học sinh không chỉ tìm được kết quả mà còn giải thích và đánh giá được tính hợp lý của kết quả đó. Lithner (2008) xem lập luận toán học là dòng tư duy được sử dụng để hình thành các khẳng định và đi đến kết luận khi giải quyết nhiệm vụ. Theo Jeannotte và Kieran (2017), lập luận gồm hai nhóm quá trình có quan hệ chặt chẽ: hình thành khẳng định toán học thông qua nhận dạng quy luật, so sánh, khái quát hóa và phán đoán; xác nhận khẳng định thông qua giải thích, biện minh và chứng minh. Quan niệm này tiếp tục được sử dụng trong các nghiên cứu gần đây về phát triển lập luận toán học (Ødegaard et al., 2026).

Khung đánh giá Toán học PISA 2022 xác định lập luận toán học, bao gồm lập luận quy nạp và diễn dịch, là nền tảng của năng lực hiểu biết toán học. Học sinh thực hiện lập luận khi phân tích tình huống, lựa chọn chiến lược, thiết lập mối liên hệ, đưa ra kết luận logic, giải thích, phê bình hoặc bác bỏ một nhận định (OECD, 2023). Vì vậy, lập luận toán học không đồng nhất với việc thực hiện phép tính hoặc tái hiện một thuật toán. Học sinh có thể tìm được đáp án đúng bằng cách áp dụng máy móc quy trình nhưng chưa chắc hiểu căn cứ toán học của quy trình đó.

Tại Việt Nam, Chương trình Giáo dục phổ thông môn Toán xác định năng lực tư duy và lập luận toán học là một trong năm thành tố cốt lõi của năng lực toán học. Ở cấp THPT, năng lực này thể hiện qua việc thực hiện các thao tác tư duy; chỉ ra căn cứ của lập luận; đưa ra kết luận hợp lý; giải thích và điều chỉnh cách thức giải quyết vấn đề (Bộ Giáo dục và Đào tạo, 2018). Nghiên cứu của Säfström et al. (2024) cũng cho thấy khó khăn lập luận của học sinh có thể xuất hiện ở các giai đoạn diễn giải thông tin, khám phá quan hệ, hình thành ý tưởng và kiểm chứng kết quả.

Từ các quan điểm trên, nghiên cứu xác định năng lực lập luận toán học của học sinh THPT gồm bốn thành phần: phân tích tình huống và xác định điều kiện; hình thành phán đoán hoặc lựa chọn chiến lược; sử dụng căn cứ để thiết lập lập luận; đánh giá, phản bác và điều chỉnh kết luận. Đây là cơ sở để xây dựng các chỉ báo và nhiệm vụ đánh giá trong nghiên cứu.

2.2. Đánh giá năng lực lập luận toán học

Đánh giá năng lực lập luận toán học là quá trình thu thập và phân tích minh chứng về cách học sinh hình thành, sử dụng và kiểm chứng các lập luận khi thực hiện nhiệm vụ toán học. Đánh giá này không chỉ xác định đáp án đúng hay sai mà còn xem xét cách học sinh phân tích dữ kiện, lựa chọn chiến lược, sử dụng căn cứ và điều chỉnh kết luận. Demosthenous et al. (2021) cho rằng đánh giá trong lớp học Toán bao gồm ba hoạt động liên hệ chặt chẽ: thiết kế nhiệm vụ để khơi gợi minh chứng, diễn giải câu trả lời của học sinh và sử dụng kết quả để hỗ trợ dạy học. Quan điểm này phù hợp với yêu cầu đánh giá vì sự tiến bộ của học sinh trong Thông tư số 22/2021/TT-BGDĐT.

Theo OECD (2023), nhiệm vụ đánh giá lập luận cần tạo cơ hội để học sinh nhận dạng và kết nối thông tin, đưa ra phán đoán, giải thích, biện minh, phê bình hoặc bác bỏ một kết luận. Vì vậy, các câu hỏi chỉ yêu cầu ghi nhớ công thức hoặc thực hiện thuật toán quen thuộc thường không cung cấp đủ minh chứng về năng lực lập luận. Nhiệm vụ phù hợp nên yêu cầu học sinh trình bày căn cứ, so sánh các phương án, tìm phản ví dụ, phát hiện sai lầm hoặc sửa chữa lời giải. Cần lưu ý rằng nhiệm vụ khó chưa chắc là nhiệm vụ đánh giá lập luận; một bài toán nhiều phép tính vẫn có thể được giải bằng quy trình máy móc.

Việc chấm điểm cần dựa trên rubric phân tích gắn với các thành phần năng lực đã xác định. Brookhart (2018) nhấn mạnh rằng rubric phải có tiêu chí phù hợp với mục đích đánh giá và mô tả rõ chất lượng thực hiện ở từng mức độ. Đối với lập luận toán học, rubric cần phân biệt giữa việc chỉ đưa ra kết luận, đưa ra căn cứ nhưng chưa đầy đủ và xây dựng được lập luận logic, chặt chẽ. Rubric cũng cần được thẩm định và kiểm tra độ thống nhất giữa những người chấm.

Trong nghiên cứu này, đánh giá năng lực lập luận toán học được thực hiện thông qua các nhiệm vụ tự luận mở. Minh chứng được thu thập theo bốn thành phần: phân tích tình huống; hình thành phán đoán hoặc chiến lược; thiết lập lập luận có căn cứ; đánh giá và điều chỉnh kết luận. Cách tiếp cận này giúp nhận diện cả kết quả đạt được và những khó khăn cụ thể trong quá trình lập luận của học sinh (Säfström et al., 2024).

2.3. Trí tuệ nhân tạo tạo sinh trong dạy học và đánh giá môn Toán

Trí tuệ nhân tạo tạo sinh là nhóm công nghệ có khả năng tạo nội dung mới dưới dạng văn bản,

hình ảnh hoặc mã lệnh dựa trên dữ liệu đã được huấn luyện. Trong dạy học Toán, các công cụ như ChatGPT, Gemini và Copilot có thể tạo bài toán, đề xuất phương pháp giải, giải thích khái niệm và cung cấp phản hồi. Tổng quan của Walkington (2025) cho thấy AI tạo sinh có tiềm năng hỗ trợ giải toán, cá nhân hóa nhiệm vụ, cung cấp hướng dẫn và hỗ trợ giáo viên lập kế hoạch dạy học.

Câu trả lời do AI tạo ra không phải lúc nào cũng chính xác. AI có thể sử dụng sai công thức, bỏ qua điều kiện, thực hiện biến đổi không hợp lệ hoặc đưa ra kết luận thiếu căn cứ. Jia et al. (2024) nhận thấy các gợi ý giải toán do AI tạo ra có chất lượng trung bình thấp hơn gợi ý của giáo viên, mặc dù người đánh giá khó phân biệt nguồn tạo ra chúng. Trong các nhiệm vụ chứng minh, Yoon et al. (2024) cũng cho thấy người học cần được hướng dẫn đánh giá có phê phán thay vì thụ động chấp nhận lập luận của AI. Ngôn ngữ trôi chảy của AI vì thế không thể được xem là bằng chứng về tính đúng đắn toán học.

AI tạo sinh đồng thời đặt ra thách thức đối với đánh giá. Học sinh có thể giao toàn bộ nhiệm vụ cho AI mà không trực tiếp thực hiện hoạt động tư duy. OECD (2026) lưu ý rằng việc AI giúp nâng cao chất lượng sản phẩm tức thời không nhất thiết tạo ra tiến bộ học tập. Vì vậy, nhiệm vụ đánh giá cần chuyển trọng tâm từ yêu cầu đáp án sang yêu cầu giải thích, kiểm chứng, so sánh và sửa chữa lập luận. Bên cạnh đó, giáo viên cần quan tâm đến tính riêng tư, sự công bằng trong tiếp cận công nghệ và tính minh bạch khi sử dụng AI (UNESCO, 2023).

Trong nghiên cứu này, sản phẩm AI được sử dụng như ngữ liệu của nhiệm vụ đánh giá, không phải công cụ làm bài thay học sinh. Tất cả học sinh được cung cấp cùng một lời giải AI và được yêu cầu xác định điều kiện, phát hiện sai lầm, đưa ra căn cứ, phản ví dụ hoặc cách sửa chữa. Cách tổ chức này giúp giảm ảnh hưởng của thiết bị và kỹ năng viết câu lệnh, qua đó tập trung đánh giá năng lực lập luận toán học.

2.4. Khung nghiên cứu đánh giá năng lực lập luận toán học trong bối cảnh AI

Nghiên cứu sử dụng tiếp cận thiết kế đánh giá dựa trên minh chứng (Evidence-Centered Design – ECD) của Mislevy và Haertel (2006). Theo tiếp cận này, nhiệm vụ đánh giá phải tạo ra những biểu hiện có thể quan sát, từ đó cho phép người nghiên cứu đưa ra nhận định về năng lực của học sinh. Khung ECD gồm ba thành phần chính: mô hình năng lực xác định nội dung cần đo; mô hình nhiệm vụ quy định đặc điểm của tình huống đánh giá; và mô hình minh chứng mô tả sản phẩm cần thu thập cùng cách diễn giải sản phẩm đó. Cách tiếp cận này phù hợp với quan điểm của Demosthenous et al. (2021), theo đó đánh giá Toán học cần bảo đảm sự liên kết giữa thiết kế nhiệm vụ, câu trả lời của học sinh và kết luận về năng lực.

Trong bối cảnh AI tạo sinh, thách thức quan trọng là bảo đảm điểm số phản ánh năng lực lập luận toán học chứ không phải khả năng sử dụng thiết bị hoặc viết câu lệnh. OECD (2026) cảnh báo rằng AI có thể nâng cao chất lượng sản phẩm tức thời nhưng không bảo đảm sự tiến bộ thực chất nếu người học giao phó hoạt động nhận thức cho công cụ. Vì vậy, nghiên cứu không cho học sinh trực tiếp sử dụng AI trong quá trình làm bài. Thay vào đó, tất cả học sinh được cung cấp cùng một lời giải AI dưới dạng văn bản để phân tích và phản biện.



Hình 1. Khung đánh giá năng lực lập luận toán học trong bối cảnh AI tạo sinh

Như vậy, tiến trình đánh giá được xác lập theo mối quan hệ: ngữ liệu AI – hoạt động phân tích của học sinh – sản phẩm lập luận – minh chứng năng lực. AI chỉ đóng vai trò tạo bối cảnh cho nhiệm vụ, không phải đối tượng được đánh giá. Khung này đồng thời đáp ứng yêu cầu sử dụng AI theo hướng lấy con người làm trung tâm, bảo đảm tính minh bạch và duy trì vai trò chủ thể của người học (UNESCO, 2023).

III. PHƯƠNG PHÁP NGHIÊN CỨU

Nghiên cứu sử dụng phương pháp nghiên cứu thiết kế giáo dục, kết hợp định tính và định lượng, tại Trường THPT Nguyễn Trãi, xã Vũ Thư, tỉnh Hưng Yên. Mẫu nghiên cứu gồm 125 học sinh thuộc các khối 10, 11 và 12. Mẫu được lựa chọn theo phương pháp phân tầng kết hợp thuận tiện: học sinh được phân theo khối lớp, sau đó lựa chọn từ các lớp có khả năng tham gia đầy đủ. Dựa trên khung Đánh giá dựa trên bằng chứng (ECD), tác giả xây dựng các nhiệm vụ yêu cầu học sinh phân tích lời giải do AI tạo sinh cung cấp, phát hiện sai sót, nêu căn cứ, đề xuất phản ví dụ và sửa chữa lập luận. Nhiệm vụ được chuyên gia giáo dục toán học thẩm định trước khi thử nghiệm. Dữ liệu được xử lí bằng thống kê mô tả và phân tích định tính; thông tin người tham gia được mã hóa và bảo mật.

IV. KẾT QUẢ NGHIÊN CỨU

4.1. Thực trạng sử dụng AI tạo sinh của học sinh

Kết quả khảo sát 125 học sinh cho thấy AI tạo sinh được sử dụng tương đối phổ biến trong học tập môn Toán. Có 109 học sinh (87,2%) đã từng sử dụng AI, trong khi 16 học sinh (12,8%) chưa từng sử dụng.

Bảng 1. Tần suất sử dụng AI tạo sinh của học sinh

Tần suất sử dụng	Số lượng (n)	Tỷ lệ (%)
Thường xuyên	39	31,2
Thỉnh thoảng	48	38,4
Hiếm khi	22	17,6
Không bao giờ	16	12,8
Tổng cộng	**125**	**100,0**

Kết quả cho thấy 69,6% học sinh sử dụng AI tạo sinh từ mức thỉnh thoảng đến thường xuyên, phản ánh mức độ tiếp cận tương đối cao của học sinh đối với công nghệ này trong học tập môn Toán. Tuy nhiên, chỉ 31,2% học sinh sử dụng thường xuyên, cho thấy AI chưa trở thành công cụ học tập ổn định đối với toàn bộ học sinh. Bên cạnh đó, 30,4% học sinh hiếm khi hoặc chưa từng sử dụng AI. Sự khác biệt này có thể liên quan đến điều kiện tiếp cận công nghệ, kỹ năng sử dụng, nhu cầu học tập và quan điểm cá nhân của học sinh.

Kết quả khảo sát tạo cơ sở thực tiễn cho việc đưa lời giải do AI tạo sinh cung cấp vào nhiệm vụ đánh giá năng lực lập luận toán học. Phần lớn học sinh đã có trải nghiệm nhất định với AI nên có khả năng tiếp nhận dạng nhiệm vụ này. Tuy nhiên, tần suất sử dụng không phản ánh chất lượng sử dụng cũng như khả năng kiểm chứng lời giải. Vì vậy, nhiệm vụ đánh giá cần tập trung vào cách học sinh phát hiện sai sót, đưa ra căn cứ, xây dựng phản ví dụ và sửa chữa lập luận, thay vì chỉ đo mức độ thường xuyên sử dụng AI.

4.2. Kết quả đánh giá năng lực lập luận toán học

Năng lực lập luận toán học của 125 học sinh được đánh giá theo bốn biểu hiện: phát hiện sai sót, đưa ra căn cứ, xây dựng phản ví dụ và sửa chữa lập luận. Mỗi biểu hiện được chấm theo rubric bốn mức.

Bảng 2. Kết quả đánh giá năng lực lập luận toán học của học sinh

Biểu hiện năng lực	Mức 1n (%)	Mức 2n (%)	Mức 3n (%)	Mức 4n (%)	M	SD
Phát hiện sai sót	17 (13,6)	38 (30,4)	48 (38,4)	22 (17,6)	2,60	0,93
Đưa ra căn cứ	24 (19,2)	42 (33,6)	43 (34,4)	16 (12,8)	2,41	0,94
Xây dựng phản ví dụ	36 (28,8)	44 (35,2)	33 (26,4)	12 (9,6)	2,17	0,95
Sửa chữa lập luận	29 (23,2)	41 (32,8)	39 (31,2)	16 (12,8)	2,34	0,97
Năng lực chung	24 (19,2)	45 (36,0)	42 (33,6)	14 (11,2)	2,37	0,92

Ghi chú: Mức 1 – chưa biểu hiện; Mức 2 – biểu hiện một phần; Mức 3 – đạt yêu cầu; Mức 4 – biểu hiện thành thạo. M: điểm trung bình; SD: độ lệch chuẩn.

Kết quả cho thấy năng lực lập luận toán học của học sinh nhìn chung chưa cao, với điểm trung bình chung đạt 2,37 trên thang bốn mức. Có 55,2% học sinh thuộc mức 1 và mức 2, trong khi tỷ lệ đạt mức 4 chỉ chiếm 11,2%. Điều này cho thấy phần lớn học sinh mới thực hiện được một số thao tác lập luận riêng lẻ, chưa hình thành được chuỗi lập luận đầy đủ và chặt chẽ.

Trong bốn biểu hiện, phát hiện sai sót đạt kết quả cao nhất (M = 2,60), còn xây dựng phản ví dụ có kết quả thấp nhất (M = 2,17). Tỷ lệ học sinh đạt mức 3–4 ở hai biểu hiện này lần lượt là 56,0% và 36,0%, chênh lệch 20 điểm phần trăm. Kết quả phản ánh việc nhận biết một bước giải bất hợp lí thường dễ hơn việc xây dựng một trường hợp cụ thể để bác bỏ lập luận. Tương tự, điểm đưa ra căn cứ và sửa chữa lập luận thấp hơn điểm phát hiện sai sót cho thấy một số học sinh có thể nhận ra lỗi nhưng chưa giải thích được nguyên nhân hoặc chưa đề xuất được cách sửa phù hợp.

4.3. Biện pháp nâng cao chất lượng đánh giá năng lực lập luận

Thứ nhất, cần chuẩn hóa lời giải do AI tạo sinh cung cấp. Giáo viên lựa chọn, kiểm tra và điều chỉnh lời giải trước khi đưa vào nhiệm vụ đánh giá, bảo đảm sai sót có chủ đích, phù hợp với kiến thức của học sinh và mục tiêu cần đo. Việc sử dụng cùng một lời giải AI giúp bảo đảm tính công bằng, tránh sự khác biệt do AI tạo ra các câu trả lời không ổn định.

Thứ hai, nhiệm vụ cần được thiết kế theo chuỗi yêu cầu: phát hiện sai sót – giải thích nguyên nhân – đưa ra căn cứ hoặc phản ví dụ – sửa chữa lập luận. Kết quả khảo sát cho thấy học sinh phát hiện lỗi tốt hơn xây dựng phản ví dụ và sửa chữa lập luận. Vì vậy, chuỗi yêu cầu này giúp thu thập đầy đủ minh chứng về quá trình suy nghĩ, thay vì chỉ kiểm tra kết quả đúng hoặc sai.

Thứ ba, cần sử dụng rubric phân tích với các tiêu chí cụ thể cho từng biểu hiện năng lực. Mỗi mức phải mô tả rõ chất lượng của căn cứ, mức độ chặt chẽ của phản ví dụ và tính hợp lí của lập luận sửa chữa. Giáo viên cần sử dụng bài làm mẫu để thống nhất cách hiểu rubric và hạn chế sự chủ quan khi chấm điểm.

Thứ tư, nên kết hợp bài làm viết với câu hỏi giải thích trực tiếp hoặc yêu cầu học sinh trình bày lại quá trình suy luận. Biện pháp này giúp kiểm chứng quyền sở hữu lập luận và hạn chế việc học sinh sử dụng AI để tạo toàn bộ câu trả lời.

Thứ năm, các nhiệm vụ cần được thử nghiệm trước khi sử dụng chính thức. Kết quả thử nghiệm được dùng để điều chỉnh ngôn ngữ, độ khó, khả năng phân hóa và thời gian thực hiện. Các biện pháp trên tập trung nâng cao độ tin cậy, tính công bằng và giá trị của việc đánh giá, chưa được xem là có hiệu quả thực nghiệm nếu chưa được kiểm chứng bằng dữ liệu độc lập.

V. KẾT LUẬN

Nghiên cứu đã xây dựng khung nhiệm vụ đánh giá năng lực lập luận toán học của học sinh THPT trong bối cảnh AI tạo sinh, trong đó lời giải do AI cung cấp được chuẩn hóa và sử dụng như ngữ liệu đánh giá. Kết quả cho thấy học sinh đã có sự tiếp cận tương đối phổ biến với AI nhưng năng lực kiểm chứng, xây dựng phản ví dụ và sửa chữa lập luận còn hạn chế. Vì vậy, nhiệm vụ đánh giá cần được thiết kế theo chuỗi hoạt động phát hiện sai sót, giải thích nguyên nhân, đưa ra căn cứ, xây dựng phản ví dụ và điều chỉnh lời giải. Việc kết hợp nhiệm vụ tự luận mở với rubric phân tích giúp thu thập minh chứng rõ ràng hơn về quá trình tư duy của học sinh. Nghiên cứu góp phần định hướng đổi mới đánh giá môn Toán, đồng thời hỗ trợ học sinh hình thành tư duy độc lập, khả năng phản biện và sử dụng AI có trách nhiệm.

TÀI LIỆU THAM KHẢO

Bộ Giáo dục và Đào tạo. (2018). Chương trình giáo dục phổ thông môn Toán (Ban hành kèm theo Thông tư số 32/2018/TT-BGDĐT ngày 26 tháng 12 năm 2018).

Bộ Giáo dục và Đào tạo. (2021). Thông tư số 22/2021/TT-BGDĐT ngày 20 tháng 7 năm 2021 quy định về đánh giá học sinh trung học cơ sở và học sinh trung học phổ thông. https://vanban.chinhphu.vn/?docid=203926&pageid=27160

Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. Frontiers in Education, 3, Article 22. https://doi.org/10.3389/feduc.2018.00022

Demosthenous, E., Christou, C., & Pitta-Pantazi, D. (2021). Mathematics classroom assessment: A framework for designing assessment tasks and interpreting students' responses. European Journal of Investigation in Health, Psychology and Education, 11(3), 1088–1106. https://doi.org/10.3390/ejihpe11030081

Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. Educational Studies in Mathematics, 96(1), 1–16. https://doi.org/10.1007/s10649-017-9761-8

Jia, J., Wang, T., & Zhang, Y. (2024). The comparison of general tips for mathematical problem solving generated by generative AI with those generated by human teachers. Asia Pacific Journal of Education, 44(1), 8–28. https://doi.org/10.1080/02188791.2023.2286920

Lithner, J. (2008). A research framework for creative and imitative reasoning. Educational Studies in Mathematics, 67(3), 255–276. https://doi.org/10.1007/s10649-007-9104-2

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. Educational Measurement: Issues and Practice, 25(4), 6–20. https://doi.org/10.1111/j.1745-3992.2006.00075.x

OECD. (2023). PISA 2022 assessment and analytical framework. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

Ødegaard, R. P., Arnesen, K. K., & Langfeldt, M. B. (2026). Tools to support learning to plan discussions that promote students' mathematical reasoning. Journal of Mathematics Teacher Education, 29, 205–234. https://doi.org/10.1007/s10857-024-09668-w

Säfström, A. I., Lithner, J., Palm, T., Palmberg, B., Sidenvall, J., Andersson, C., Boström, E., & Granberg, C. (2024). Developing a diagnostic framework for primary and secondary students' reasoning difficulties during mathematical problem solving. Educational Studies in Mathematics, 115, 125–149. https://doi.org/10.1007/s10649-023-10278-1

UNESCO. (2023). Guidance for generative AI in education and research. https://doi.org/10.54675/EWZM9535