



DOI: <https://doi.org/10.52714/dthu.14.04S.2025.1585>

THIẾT KẾ VÀ ĐÁNH GIÁ THANG ĐO NĂNG LỰC GIẢI QUYẾT VẤN ĐỀ TOÁN HỌC CỦA HỌC SINH LỚP 6 Ở NỘI DUNG “GIÁ TRỊ PHÂN SỐ CỦA MỘT SỐ”

Phạm Văn Bản^{1*}, Bùi Phương Uyên² và Nguyễn Thị Thanh Loan³

¹*Khoa Sau đại học, Đại học Cần Thơ, Việt Nam*

²*Trường Sư phạm, Đại học Cần Thơ, Việt Nam*

³*Trường Phổ thông Thực hành Sư phạm, Trường Đại học An Giang, ĐHQG-HCM, Việt Nam*

**Tác giả liên hệ, Email: banp2224001@gstudent.ctu.edu.vn*

Lịch sử bài báo

Ngày nhận: 21/5/2025; Ngày nhận chỉnh sửa: 16/7/2025; Ngày duyệt đăng: 22/7/2025

Tóm tắt

Dạy học phát triển năng lực, trong đó có năng lực giải quyết vấn đề toán học, là một xu thế tất yếu trong giáo dục. Việc đánh giá chính xác năng lực của học sinh sẽ là cơ sở quan trọng để đề xuất phương pháp dạy học phù hợp cũng như đánh giá hiệu quả của các phương pháp đó. Nghiên cứu này đã thiết kế một rubric dành cho đánh giá các bài kiểm tra viết về năng lực giải quyết vấn đề toán học. Thông qua đánh giá của 66 giáo viên và thực nghiệm trên 132 học sinh, nghiên cứu đã đề xuất một rubric tốt cho việc đánh giá, đồng thời cũng đưa ra những kiến nghị cho việc thiết kế các rubric tương tự trong tương lai.

Từ khóa: *đánh giá, giải quyết vấn đề toán học, lớp 6, năng lực, rubric.*

Trích dẫn: Phạm, V. B., Bùi, P. U., & Nguyễn, T. T. L. (2025). Thiết kế và đánh giá thang đo năng lực giải quyết vấn đề toán học của học sinh lớp 6 ở nội dung “Giá trị phân số của một số”. *Tạp chí Khoa học Đại học Đồng Tháp*, 14(04S), 254-268. <https://doi.org/10.52714/dthu.14.04S.2025.1585>

Copyright © 2025 The author(s). This work is licensed under a CC BY-NC 4.0 License.

DESIGNING AND EVALUATING A SCALE OF MATHEMATICAL PROBLEM-SOLVING ABILITY OF GRADE 6 STUDENTS ON THE CONTENT "FRACRATIONAL VALUE OF A NUMBER"

Pham Van Ban^{1*}, Bui Phuong Uyen², Nguyen Thi Thanh Loan³

¹*Graduate School, Can Tho University, Vietnam*

²*School of Education, Can Tho University, Vietnam*

³*Pedagogical Practice School, An Giang University, VNU-HCM, Vietnam*

**Corresponding author, Email: banp2224001@gstudent.ctu.edu.vn*

Article history

Received: 21/5/2025; Received in revised form: 16/7/2025; Accepted: 22/7/2025

Abstract

Developing competencies, including mathematical problem-solving ones, is an inevitable trend in education. Assessing students' competencies accurately will be an important basis for proposing appropriate teaching methods as well as evaluating the effectiveness of those methods. This study designed a rubric for assessing written tests on mathematical problem-solving competencies. Through the assessment of 66 teachers and experiments on 132 students, the study proposed a good rubric for assessment, and also made recommendations for designing similar rubrics in the future.

Keywords: *assessment, mathematical problem-solving, grade 6, competencies, rubric.*

1. Giới thiệu

Đánh giá là một phần quan trọng của quá trình dạy học. Đánh giá được các nhà giáo dục định nghĩa là một sự nỗ lực liên tục và có hệ thống để thu thập thông tin về kết quả học tập của người học (Darmayanti, Syaifuddin, & cs, 2022; Hung & Wang, 2021; Leacock & Nesbit, 2007). Kết quả kiểm tra đánh giá sẽ cung cấp một thông tin tổng quan về thành tích học tập của học sinh trong việc hiểu nội dung của môn học hoặc một số năng lực được mong đợi học sinh sẽ đạt được trong quá trình học tập (Choirudin & cs, 2022; Nurtanto & cs, 2021; Otgonbaatar, 2021). Các nguyên tắc chung của việc đánh giá bao gồm: (1) Dựa trên kỹ năng; (2) Hợp lệ; (3) Phù hợp; (4) Khách quan; (5) Bền vững; (6) Toàn diện; (7) Có tính mở và (8) Phương tiện (Özer & cs, 2020; Palinussa & cs, 2021; Suwarno & Aeni, 2021).

Việc đánh giá thường được tiến hành thông qua các bài kiểm tra. Có nhiều loại bài kiểm tra khác nhau như kiểm tra viết, vấn đáp, thực hành. Đối với bài kiểm tra viết, nó được chia thành hai hình thức là trắc nghiệm khách quan và tự luận. Khác với bài kiểm tra trắc nghiệm khách quan, việc đánh giá chỉ dựa trên kết quả làm bài của học sinh, việc đánh giá bài tự luận phức thuộc vào tính chủ quan của người chấm. Asmana (2018) đã cho biết giáo viên thường sử dụng trực giác cá nhân khi đánh giá mà không có hướng dẫn cụ thể. Điều này làm cho việc đánh giá kết quả học tập hoặc năng lực của học sinh thường trở nên không hợp lệ.

Khi đánh giá năng lực giải quyết vấn đề toán học của học sinh, hầu hết giáo viên đều sử dụng các bài kiểm tra tự luận, đây là một hình thức đánh giá phù hợp (Anisah & Lastuti, 2018; Darmayanti, Sugianto, & cs, 2022; Zulfa & cs, 2019). Tuy vậy, tính chủ quan của người đánh giá đôi khi làm cho kết quả này bị thiên vị và không hợp lệ. Việc này có thể khắc phục nếu sử dụng một thang đo thích hợp. Rubric là một hướng dẫn đánh giá được giáo viên sử dụng để mô tả các tiêu chí và mức độ của kết quả học tập của học sinh (Cahyadi & cs, 2023), nó cũng được thiết kế như một hướng dẫn để giáo viên đánh giá chính xác và khách quan hơn kết quả của học sinh (Jonan, 2020; Nurhaifa & cs, 2020). Đối với học sinh, rubric thể hiện kỳ vọng học tập mà từ đó học sinh có động lực học tập hơn (Suwarno & Aeni, 2021).

1.1. Năng lực giải quyết vấn đề toán học và các thành tố

Năng lực giải quyết vấn đề toán học

Có nhiều quan niệm khác về năng lực giải quyết vấn đề toán học, nhưng hầu hết các nhà nghiên cứu đều thống nhất rằng năng lực giải quyết vấn đề toán học bao hàm các tính chất: là thuộc tính cá nhân, cho phép vận dụng những gì đã có của bản thân (kiến thức, kỹ năng, kinh nghiệm) với mục tiêu là giải quyết vấn đề trong toán học hoặc vấn đề thực tế cần vận dụng kiến thức toán. (Hoa & Dương, 2024; Nguyễn & cs, 2023; Võ & Tô, 2023).

Các thành tố và biểu hiện của năng lực giải quyết vấn đề toán học

Do năng lực giải quyết vấn đề toán học là khả năng giải quyết có hiệu quả một vấn đề toán học nào đó, trên cơ sở vận dụng các tri thức, kinh nghiệm, kỹ năng đã có nên ở các cấp học khác nhau, khi học sinh đang có trình độ về kiến thức và kỹ năng khác nhau, yêu cầu cần đạt của năng lực giải quyết vấn đề toán học sẽ khác nhau. Theo Bộ Giáo dục và Đào tạo (2018), yêu cầu cần đạt của năng lực này ở các cấp học như trong Bảng 1.

Bảng 1. Thành phần và các yêu cầu cần đạt của năng lực giải quyết vấn đề toán học

Thành phần năng lực	Cấp tiểu học	Cấp trung học cơ sở	Cấp trung học phổ thông
---------------------	--------------	---------------------	-------------------------

Nhận biết, phát hiện được vấn đề cần giải quyết bằng toán học	Nhận biết được vấn đề cần giải quyết và nêu được thành câu hỏi	Phát hiện được vấn đề cần giải quyết	Xác định được tình huống có vấn đề, thu thập, sắp xếp, giải thích và đánh giá được độ tin cậy của thông tin, chia sẻ sự am hiểu vấn đề với người khác
Lựa chọn, đề xuất được cách thức, giải pháp giải quyết vấn đề	Nêu được cách thức giải quyết vấn đề	Xác định được cách thức, giải pháp giải quyết vấn đề	Lựa chọn và thiết lập được cách thức, quy trình giải quyết vấn đề
Sử dụng được các kiến thức, kỹ năng toán học tương thích (bao gồm các công cụ và thuật toán) để giải quyết vấn đề đặt ra	Thực hiện và trình bày được cách thức giải quyết vấn đề ở mức độ đơn giản	Sử dụng được các kiến thức, kỹ năng toán học tương thích để giải quyết vấn đề	Thực hiện và trình bày được giải pháp giải quyết vấn đề
Đánh giá được giải pháp đề ra và khái quát hóa được cho vấn đề tương tự	Kiểm tra được giải pháp đã thực hiện	Giải thích được giải pháp đã thực hiện	Đánh giá được giải pháp đã thực hiện, phản ánh được giá trị của giải pháp, khái quát hóa được cho vấn đề tương tự.

Theo chương trình giáo dục môn Toán năm 2018, năng lực giải quyết vấn đề toán học là một năng lực đặc thù quan trọng của môn Toán. Có nhiều nghiên cứu trong và ngoài nước đã thực hiện thiết kế các thang đo năng lực giải quyết vấn đề toán học, chúng tôi đã tiến hành phân tích, so sánh và đánh giá các thang đo đó ở Bảng 2.

Bảng 2. So sánh một số thang đo năng lực GQVĐTH

Tác giả, năm	Đối tượng học sinh	Nội dung	Ưu điểm	Hạn chế
Cục đánh giá học sinh Chicago (trích theo Charles & cs. (2008))	Học sinh nói chung, không phân cấp và nội dung	Toán học nói chung	Là thang đo được xây dựng sớm dựa trên cơ sở mô hình của G. Polya, là định hướng cho các nghiên cứu về sau	Không thể hiện rõ yêu cầu cần đạt cho năng lực giải quyết vấn đề toán học

Wu (2003)	Học sinh trung học	Toán học nói chung	Phù hợp với các bài toán liên quan thực tế	Cần dữ liệu lớn và áp dụng lý thuyết IRT, khó áp dụng ở Việt Nam
Gunawardena (2010)	Sinh viên cao đẳng	Toán học nói chung	Phù hợp với các bài toán thuần túy hơn là các bài toán có yếu tố thực tiễn, phù hợp học sinh THPT và sinh viên	Thang đo chưa phù hợp với học sinh cấp THCS
Phan (2014)	Học sinh lớp 11	Toán học nói chung	Thể hiện mối liên giữa các tiêu chí và thang đo tổng thể, giúp đưa ra kết luận về năng lực một cách thuận tiện	Chưa thực sự tương thích với Chương trình năm 2018, các mức độ chưa được định tính và định lượng một cách rõ ràng
Salazar-Torres & cs. (2021)	Học sinh THPT	Toán thực tiễn có yếu tố vật lý	Thang đo từng tiêu chí chi tiết, có đánh giá chỉ số năng lực giải quyết vấn đề toán học	Thang đo công kênh, cần đánh giá quá nhiều chi tiết, không phù hợp cho các bài toán của lớp nhỏ
Nguyễn & cs. (2023)	Học sinh lớp 10	Hệ thức lượng trong tam giác	Chi tiết về các tiêu chí và mức độ, thuận tiện cho áp dụng trong thực tế	Chưa hoàn toàn tương thích với các yêu cầu cần đạt trong Chương trình năm 2018

Đối với học sinh lớp 6, hầu hết các yêu cầu cần đạt ở môn Toán đều ở mức độ nhận biết và tính toán (Bộ Giáo dục và Đào tạo, 2018). Do vậy, các biểu hiện của năng lực giải quyết vấn đề toán học của học sinh lớp 6 cũng khác biệt với học sinh các lớp khác. Từ đó, cần thiết phải thiết kế một rubric để đánh giá bài tự luận cho năng lực giải quyết vấn đề toán học cho học sinh lớp 6.

1.2. Mục tiêu nghiên cứu

Từ những mô tả bên trên, mục tiêu của nghiên cứu này là xây dựng một tiêu chuẩn để đánh giá các bài kiểm tra tự luận về năng lực giải quyết vấn đề toán học cho học sinh lớp 6. Ngoài ra, nghiên cứu cũng đề xuất trọng số của các tiêu chí trong rubric này để phù hợp với thực tế giảng dạy cũng như giúp giáo viên và học sinh tập trung nhiều hơn vào những tiêu chí quan trọng. Sau khi nghiên cứu Chương trình giáo dục môn Toán năm 2018, chúng tôi đồng ý với tác giả Nguyễn (2025) rằng chủ đề “Phân số” và nội dung “Giá trị phân số của một số” có nhiều cơ hội để phát triển kỹ năng phân tích, tổng hợp, lập luận hợp lý khi giải quyết vấn đề toán học, từ đó, chúng tôi chọn nội dung “Giá trị phân số của một số” trong chương trình Toán 6 để thiết kế rubric và tiến hành thực nghiệm.

1.3. Câu hỏi nghiên cứu

Để thực hiện được mục tiêu nghiên cứu trên, chúng tôi sẽ trả lời cho các câu hỏi nghiên cứu sau:

RQ1: Rubric đánh giá bài kiểm tra tự luận về năng lực giải quyết vấn đề toán học với nội dung “Giá trị phân số của một số” có những tiêu chí gì và trọng số các tiêu chí ra sao?

RQ2: Kết quả đánh giá mức độ phù hợp của rubric như thế nào?

2. Phương pháp nghiên cứu

2.1. Thiết kế thang đo

Từ các yêu cầu cần đạt của năng lực giải quyết vấn đề toán học ở cấp trung học cơ sở (Bộ Giáo dục và Đào tạo, 2018), tham khảo các thang đánh giá của Cục Đánh giá học sinh của các trường công lập tại Chicago, Hoa Kỳ (1987) (trích theo Charles & cs. (2008)), Wu (2003), Gunawardena (2010), Phan (2014), Salazar-Torres & cs. (2021), Nguyễn & cs. (2023)..., trong nghiên cứu này, chúng tôi thiết kế một thang đo dành cho bài kiểm tra nội dung “Giá trị phân số của một số” của học sinh lớp 6. Với thang đo này, giáo viên có một công cụ để đánh giá một cách khách quan bài làm của học sinh, và học sinh có một bảng định hướng để tự học.

Cấu trúc thang đo: Thang đo được xây dựng chủ yếu dựa trên bốn yêu cầu cần đạt ở cấp trung học cơ sở của năng lực giải quyết vấn đề toán học được quy định trong Chương trình giáo dục môn Toán năm 2018.

Bảng 3. Cấu trúc thang đo năng lực giải quyết vấn đề toán học cho học sinh lớp 6

Tiêu chí (Thành phần năng lực)	Yêu cầu cần đạt	Trọng số (%)
(TC1) Nhận biết, phát hiện được vấn đề cần giải quyết bằng toán học	Phát hiện được vấn đề cần giải quyết	25
(TC2) Lựa chọn, đề xuất được cách thức, giải pháp giải quyết vấn đề	Xác định được cách thức, giải pháp giải quyết vấn đề	25
(TC3) Sử dụng được các kiến thức, kỹ năng toán học tương thích (bao gồm các công cụ và thuật toán) để giải quyết vấn đề đặt ra	Sử dụng được các kiến thức, kỹ năng toán học tương thích để giải quyết vấn đề	40
(TC4) Đánh giá được giải pháp đề ra và khái quát hóa được cho vấn đề tương tự	Giải thích được giải pháp đã thực hiện	10

Rubric đánh giá từng tiêu chí

Bảng 4. Rubric đánh giá bài viết về chủ đề "Giá trị phân số của một số"

Tiêu chí	Mức 3 (Tốt)	Mức 2 (Khá)	Mức 1 (TB)	Mức 0 (Yếu)
TC1. Phát hiện vấn đề cần giải quyết	Xác định đúng giá trị cần tính và các giá trị trung gian	Xác định đúng giá trị cần tính nhưng hiểu sai các giá trị trung gian	Hiểu sai giá trị cần tính, hiểu đúng một số giá trị trung gian	Không hiểu yêu cầu của bài và các giá trị có trong đề bài

TC2. Xác định cách giải quyết vấn đề	Các bước tính toán đầy đủ và hợp lý	Các bước tính toán hợp lý nhưng chưa đầy đủ	Các bước tính toán chưa hợp lý nhưng khá đầy đủ	Các bước lộn xộn hoặc không thể hiện được các bước cần tính
TC3. Sử dụng kiến thức toán học phù hợp	Thực hiện tính toán đầy đủ và chính xác	Tính đúng giá trị cần tính, có sai sót nhỏ ở các bước trung gian	Tính đúng ở một số bước trung gian nhưng sai giá trị cần tính	Tính sai những giá trị trung gian quan trọng nên sai kết quả hoặc không làm
TC4. Giải thích giải pháp đã thực hiện	Ở mỗi bước có ghi lời giải đầy đủ	Có lời giải ở giá trị cần tính, thiếu lời giải ở một số bước trung gian	Có lời giải ở một số bước trung gian, thiếu lời giải ở các bước chính	Không có giải thích hoặc chỉ viết kết quả

Chúng tôi đề xuất một bài kiểm tra viết ngắn để học sinh làm, chấm điểm theo rubric và từ đó đánh giá tính hợp lý của rubric. Đề bài: “*Bạn Mai đọc một quyển sách 120 trang trong 4 ngày. Ngày đầu tiên Mai đọc được 1/3 quyển sách. Ngày thứ hai Mai đọc được 1/4 số trang còn lại. Ngày cuối cùng Mai đọc được số trang bằng 3/2 số trang của ngày thứ hai. Hỏi ngày thứ ba Mai đọc được bao nhiêu trang?*”

2.2. Đánh giá nội dung thang đo

Thiết kế nghiên cứu: Chúng tôi tiến hành khảo sát ý kiến của các chuyên gia là giáo viên toán đang trực tiếp giảng dạy ở các trường trung học cơ sở về thang đo thông qua một bảng hỏi. Bảng hỏi gồm hai phần, phần thứ nhất đánh giá chung về mức độ phù hợp của rubric và phần thứ hai đánh giá mức độ phù hợp của từng tiêu chí. Ở phần thứ nhất, chúng tôi khảo sát ý kiến thông qua các câu hỏi (Bùi & cs, 2023):

CH1: Các tiêu chí đánh giá phản ánh đầy đủ nội dung và năng lực giải quyết vấn đề toán học

CH2: Các mức độ khác nhau được đặt tên và giá trị điểm số phù hợp

CH3: Các thông tin mô tả rõ ràng, thể hiện theo một chuỗi liên kết và đảm bảo cho sự phát triển của học sinh

CH4: Ngôn ngữ rõ ràng, dễ hiểu với học sinh

CH5: Dễ sử dụng với giáo viên

CH6: Học sinh có thể xác định dễ dàng những kỹ năng phát triển cần thiết

Các câu hỏi này được thiết kế theo thang Likert 4 bậc là 1 = Rất không đồng ý, 2 = Không đồng ý, 3 = Đồng ý, 4 = Rất đồng ý

Ở phần thứ hai, chúng tôi khảo sát ý kiến về từng tiêu chí trong rubric theo thang Likert 4 bậc là 1 = Không phù hợp, 2 = Hơi phù hợp, 3 = Phù hợp, 4 = Rất phù hợp.

Trong quá trình trao đổi với giáo viên, chúng tôi nhận thấy nếu chia quá nhiều mức độ đánh giá, mỗi mức độ sẽ có biên độ khá hẹp, dẫn đến cùng quan điểm đánh giá như nhau, nhưng các giáo viên lại có lựa chọn mức độ khác nhau. Ngoài ra, để thuận tiện cho việc khảo sát ý kiến mở rộng của giáo viên, chúng tôi dùng thang Likert 4 bậc để ở mỗi câu hỏi, nếu giáo viên chọn giá trị 1 hoặc 2 (hai mức thấp) thì nhờ giáo viên góp ý về nội dung tương ứng để chỉnh sửa.

Mẫu khảo sát: Dựa trên số biến quan sát của bảng hỏi, chúng tôi kỳ vọng khảo sát trên mẫu có kích thước nhỏ nhất là 50 giáo viên dạy toán từ các trường trung học cơ sở (Bryman, 2016). Mẫu khảo sát được thực hiện theo hình thức mẫu ngẫu nhiên thông qua ứng dụng Google Form. Việc khảo sát được tiến hành ngẫu nhiên tuân thủ các nguyên tắc về điều tra xã hội học do Trần (2008) đề xuất.

Thời gian khảo sát: Chúng tôi tiến hành khảo sát trực tuyến từ ngày 27/02/2025 đến ngày 06/03/2025, liên kết đến trang khảo sát được chúng tôi gửi đến giáo viên thông qua các cộng đồng trực tuyến trên các mạng xã hội Facebook, Zalo.

2.3. Đánh giá thực nghiệm thang đo

Thiết kế nghiên cứu: Chúng tôi đã lấy bài toán được đề xuất ở mục bên trên và cho học sinh lớp 6 làm trong thời gian 15 phút và thu lại bài làm để tiến hành chấm điểm theo từng tiêu chí. Việc chấm điểm được thực hiện qua 3 giai đoạn (Dawson, 2017; DeVellis, 2017; Lamm & cs, 2020; Rezaei & Lovorn, 2010):

- Giai đoạn 1: Hướng dẫn và nhờ một giáo viên đang trực tiếp giảng dạy ở các lớp 6 này (GV1) chấm điểm theo từng tiêu chí của rubric, kết quả được đánh dấu vào mỗi bảng rubric được kẹp cùng bài làm của học sinh, giáo viên không tác động gì vào bài làm của học sinh.

- Giai đoạn 2: Hướng dẫn và nhờ một giáo viên khác có dạy lớp 6 (GV2) chấm lại 20% bài làm của học sinh. Các bài này được chọn một cách ngẫu nhiên.

- Giai đoạn 3: Sau khi hoàn thành giai đoạn 1 hai tuần, chúng tôi nhờ GV1 chấm lại 20% bài làm được chọn ngẫu nhiên của học sinh.

Mẫu nghiên cứu: Chúng tôi tiến hành thực nghiệm trên 132 học sinh lớp 6 của ba lớp ở trường Phổ thông Thực hành Sư phạm (Trường Đại học An Giang). Các lớp này đều do cùng GV1 trực tiếp giảng dạy trên lớp. Trong quá trình chọn mẫu, chúng tôi đã chọn mẫu thuận tiện theo các bước: Trao đổi và chọn giáo viên thực nghiệm trên tinh thần tự nguyện, sau đó chọn GV1 là giáo viên dạy nhiều lớp hơn để có mẫu thực nghiệm lớn hơn, giáo viên còn lại sẽ là GV2.

Thời gian thực nghiệm: Chúng tôi đã tiến hành thực nghiệm vào ngày 07/03/2025 bằng cách chia học sinh thành 3 phòng để làm bài kiểm tra cùng một lúc, thời gian làm bài là 15 phút.

Xử lý số liệu: Kết quả bài làm được mã hóa và xử lý bằng các phần mềm Microsoft Excel và Jamovi để biểu diễn trực quan các kết quả, tính toán các hệ số Cronbach's Alpha, thực hiện các kiểm định T-Test, ANOVA, Chi-bình phương và hệ số tương quan (The Jamovi Project, 2024).

3. Kết quả và phân tích

3.1. Đánh giá nội dung

Chúng tôi đã tiến hành khảo sát ý kiến của 66 giáo viên trung học cơ sở trong thời gian từ ngày 27/02/2025 đến ngày 06/03/2025. Kết quả khảo sát cụ thể như sau:

Bảng 5. Kết quả đánh giá mức độ phù hợp của rubric

Câu hỏi	CH1	CH2	CH3	CH4	CH5	CH6
Trung bình	3,4697	3,3939	3,4545	3,4848	3,4848	3,4394

Trung bình kết quả đánh giá của các câu hỏi nằm trong khoảng 3,3939 đến 3,4848. Tỷ lệ giáo viên có đánh giá mức phù hợp và rất phù hợp chiếm 97,73%. Kết quả đánh giá này cho thấy hầu hết giáo viên đều đánh giá rubric phù hợp để đánh giá bài làm của học sinh về năng lực giải quyết vấn đề toán học.

Mức độ phù hợp của mỗi tiêu chí trong rubric được thể hiện ở Bảng 6.

Bảng 6. Kết quả đánh giá mức độ phù hợp của mỗi tiêu chí

Tiêu chí	TC1	TC2	TC3	TC4
Trung bình	3,3788	3,3939	3,4848	3,4091

Tỷ lệ giáo viên đánh giá các tiêu chí mức phù hợp và rất phù hợp là 95,45%, trung bình của mỗi tiêu chí nằm trong khoảng 3,3939 đến 3,4848. Kết quả này cho thấy các hầu hết giáo viên đều đánh giá cách mô tả các tiêu chí là phù hợp để đánh giá các bài kiểm tra về năng lực giải quyết vấn đề toán học của học sinh.

Do trung bình đánh giá của các tiêu chí đều cao hơn 2,5 nên chúng tôi không thay đổi mô tả các tiêu chí và các mức độ. Sau bước này, chúng tôi sẽ tiến hành đánh giá bài kiểm tra của học sinh để đánh giá tính phù hợp thực nghiệm của rubric.

3.2. Kiểm tra tính nhất quán bên trong

Ở Giai đoạn 1, sau khi cho 132 học sinh làm bài kiểm tra, chúng tôi đã hướng dẫn giáo viên GV1 cách thức sử dụng rubric để chấm các bài kiểm tra này. Các bài kiểm tra được chấm bằng cách đánh dấu vào mức độ tương ứng của mỗi tiêu chí trong rubric, sau đó chuyển thành điểm của từng tiêu chí trong danh sách. Với mỗi bài làm, chúng tôi tính ra điểm tổng dựa trên trọng số của từng tiêu chí như sau:

$$TB = TC1 * 25\% + TC2 * 25\% + TC3 * 40\% + TC4 * 10\%.$$

Dựa trên điểm trung bình của từng bài làm, chúng tôi phân hạng các bài làm theo tiêu chí: 27% bài (36 bài) có điểm trung bình cao nhất được xếp loại Giỏi, 27% bài có điểm trung bình thấp nhất được xếp loại Trung bình, còn lại được xếp loại Khá. Dữ liệu thu được trình bày như trong bảng sau:

ID	TC1	TC2	TC3	TC4	TB	Group
ST001	2	3	3	3	2,75	Giỏi
ST002	2	3	3	2	2,65	Giỏi
...						
ST006	0	0	0	0	0	Trung bình
ST007	2	2	1	0	1,40	Khá

ST008	1	0	1	0	0,65	Trung bình
...						
ST132	0	0	0	0	0	Trung bình

Nhằm đánh giá tính nhất quán của các tiêu chí trong rubric, một phân tích về hệ số Cronbach's Alpha giữa các tiêu chí TC1, TC2, TC3 và TC4 đã được tiến hành, kết quả thu được như trong Bảng 7.

Bảng 7. Kết quả tính hệ số Cronbach's Alpha

Scale Reliability Statistics		
	Cronbach's α	McDonald's ω
scale	0,931	0,934

Cả hai giá trị Cronbach's Alpha và McDonald's Omega đều đạt giá trị rất cao và gần bằng nhau (lần lượt là 0,931 và 0,934), điều này cho thấy các tiêu chí trong rubric có độ nhất quán cao, đo lường cùng một năng lực một cách ổn định. (Cohen & cs, 2017).

Phân tích tương quan Spearman được thực hiện để kiểm tra mối quan hệ giữa các tiêu chí trong rubric. Kết quả cho thấy tất cả các tiêu chí có tương quan dương, mạnh và có ý nghĩa thống kê ($p < 0,001$). Mối tương quan cao nhất được quan sát giữa TC2 và TC3 ($r_s = 0,862, p < 0,001$), cho thấy hai tiêu chí này có mức độ liên kết rất chặt chẽ. TC1 có mối tương quan cao với TC2 ($r_s = 0,826, p < 0,001$), nhưng mức tương quan với TC3 ($r_s = 0,684, p < 0,001$) và TC4 ($r_s = 0,636, p < 0,001$) thấp hơn. TC4 có tương quan cao với TC3 ($r_s = 0,865, p < 0,001$), cho thấy hai tiêu chí này có thể đo lường cùng một khía cạnh (Cohen & cs, 2017).

3.3. Kiểm tra tính hợp lệ thực nghiệm

Phương pháp trích xuất phần dư tối thiểu (minimum residual extraction) kết hợp với phép quay oblimin rotation đã được sử dụng. Kết quả cho thấy tất cả các tiêu chí đều tải mạnh vào một nhân tố duy nhất, với hệ số tải dao động từ 0,786 đến 0,957. Tiêu chí TC2 (0,957) và TC3 (0,926) có hệ số tải cao nhất, đồng thời có giá trị uniqueness rất thấp (0,084 và 0,143), cho thấy chúng có thể đang đo lường cùng một nội dung. Tiêu chí TC1 (0,786) và TC4 (0,855) có giá trị uniqueness cao hơn (0,382 và 0,269), thể hiện sự khác biệt nhất định nhưng vẫn thuộc cùng một nhân tố.

Kết quả này cho thấy rubric đo lường một khía cạnh duy nhất, đồng thời có tính nhất quán nội tại tốt (Cronbach's $\alpha = 0,931$). Tuy nhiên, do mức độ tương đồng rất cao giữa một số tiêu chí (TC2–TC3), có thể xem xét loại bỏ một tiêu chí nếu mục tiêu là đơn giản hóa rubric (Cohen & cs, 2017).

Một phân tích nhân tố xác nhận (CFA) đã được thực hiện để kiểm tra cấu trúc tiềm ẩn của bộ tiêu chí đánh giá. Kết quả cho thấy tất cả các tiêu chí đều có hệ số tải nhân tố cao (TC1 = 0,798, TC2 = 1,007, TC3 = 1,031, TC4 = 1,042) và có ý nghĩa thống kê ($p < 0,001$). Điều

này có nghĩa là các tiêu chí này có mối liên hệ chặt chẽ với nhân tố chung, nghĩa là chúng đo lường một khái niệm chung (Bryman, 2016).

Do sự tương đồng cao giữa TC2 và TC3, trong lần thực nghiệm tiếp theo, chúng tôi dự kiến kết hợp 2 tiêu chí này thành một tiêu chí trong rubric, khi đó, thang đo mới dự kiến sẽ thay đổi như Bảng 8.

Bảng 8. Rubric đánh giá dự kiến cho lần thực nghiệm tiếp theo

Tiêu chí	Mức 3 (Tốt)	Mức 2 (Khá)	Mức 1 (TB)	Mức 0 (Yếu)	Trọng số
TC1. Phát hiện vấn đề cần giải quyết	Xác định đúng giá trị cần tính và các giá trị trung gian	Xác định đúng giá trị cần tính nhưng hiểu sai các giá trị trung gian	Hiểu sai giá trị cần tính, hiểu đúng một số giá trị trung gian	Không hiểu yêu cầu của bài và các giá trị có trong đề bài	25%
TC2. Xác định phương pháp và thực hiện giải quyết vấn đề	Phương pháp hợp lý, đầy đủ và thực hiện chính xác	Phương pháp hợp lý và tính toán có sai sót nhỏ ở các bước trung gian	Phương pháp đầy đủ nhưng chưa hợp lý hoặc chỉ tính đúng ở các bước trung gian, sai giá trị cần tính	Không thể hiện các bước rõ ràng hoặc tính sai các giá trị trung gian quan trọng hoặc không làm	60%
TC3. Giải thích giải pháp đã thực hiện	Ở mỗi bước có ghi lời giải đầy đủ	Có lời giải ở giá trị cần tính, thiếu lời giải ở một số bước trung gian	Có lời giải ở một số bước trung gian, thiếu lời giải ở các bước chính	Không có giải thích hoặc chỉ viết kết quả	15%

Ở tiêu chí TC2, chúng tôi kết hợp các liên từ “và”, “hoặc” trong các mức độ để trong quá trình đánh giá, giáo viên có thể xác định rõ hơn thành phần và năng lực mà học sinh thực sự đạt được, để từ đó chúng tôi tiếp tục cân nhắc việc giữa các tiêu chí TC2 và TC3 như phân bản đầu tiên hay gộp lại như phiên bản tiếp theo.

Kết quả phân tích ANOVA và hậu kiểm

Để kiểm tra tính phân biệt của các tiêu chí trong rubric, phân tích phương sai một yếu tố (One-Way ANOVA) với phương pháp Welch đã được thực hiện trên bốn tiêu chí (TC1, TC2, TC3, TC4) giữa ba nhóm trình độ (Giỏi, Khá, Trung bình). Kết quả phân tích ANOVA cho thấy có sự khác biệt có ý nghĩa thống kê giữa các nhóm trình độ trên tất cả các tiêu chí:

- **TC1:** $F(2; 78, 4) = 160, p < 0,001$
- **TC2:** $F(2; 78, 7) = 509, p < 0,001$
- **TC3:** $F(2; 78, 0) = 419, p < 0,001$
- **TC4:** $F(2; 84, 3) = 517, p < 0,001$

3.4. Kiểm tra tính ổn định

Ở giai đoạn 2, chúng tôi đã hướng dẫn và nhờ giáo viên GV2 chấm 20% số bài (26 bài) được chọn ngẫu nhiên. Để đánh giá mức độ đồng thuận giữa hai giám khảo, hệ số Kappa của Light và Fleiss đã được tính toán cho bốn tiêu chí (TC1, TC2, TC3, và TC4). Kết quả cho thấy mức độ đồng thuận dao động từ mạnh đến hoàn hảo.

Cụ thể, hệ số Kappa của Light cho các tiêu chí là: TC1 = 0,779, TC2 = 0,835, TC3 = 1,00, và TC4 = 0,783. Các giá trị này đều có ý nghĩa thống kê ($Z > 6,22, p < 0,001$). Hệ số Fleiss Kappa cũng cho kết quả tương tự (TC1 = 0,778, TC2 = 0,835, TC3 = 1,00, TC4 = 0,782). Điều này cho thấy mức độ nhất quán cao giữa hai giám khảo khi đánh giá các tiêu chí.

Phân tích Kappa theo từng hạng mục cho thấy mức độ đồng thuận cao hơn ở các mức điểm cao hơn. Ở tiêu chí TC1, hạng mục 0 có Kappa thấp nhất ($K = 0,336, Z = 1,71, p = 0,087$), cho thấy mức độ đồng thuận không đạt ý nghĩa thống kê. Trong khi đó, mức 3 có mức đồng thuận hoàn hảo ($K = 1,00, Z = 5,10, p < 0,001$). Xu hướng tương tự cũng được quan sát ở TC4, với Kappa thấp nhất ở mức 0 ($K = 0,623, Z = 3,18, P = 0,001$) và cao nhất ở mức 3 ($K = 0,92, Z = 4,69, p < 0,001$).

Nhìn chung, kết quả chỉ ra rằng mức độ đồng thuận giữa hai giám khảo là rất tốt, đặc biệt là ở tiêu chí TC3 (Kappa = 1,00, đồng thuận hoàn hảo). Tuy nhiên, độ tin cậy thấp hơn ở các mức điểm thấp hơn, cho thấy có thể cần thêm hướng dẫn hoặc đào tạo để đảm bảo sự nhất quán cao hơn ở tất cả các mức đánh giá (Cohen & cs, 2017).

Sau hai tuần, chúng tôi tiến hành Giai đoạn 3, chọn ngẫu nhiên 20% số bài (26 bài) để nhờ GV1 chấm lần 2 để kiểm tra tính ổn định. Để đánh giá độ tin cậy test-retest giữa hai lần chấm, hệ số tương quan Spearman (ρ) đã được sử dụng. Kết quả cho thấy tất cả các tiêu chí đều có mức tương quan cao và có ý nghĩa thống kê. Cụ thể:

Tiêu chí 1 (TC1): Hệ số tương quan Spearman rất cao, $\rho = 0,94, df = 24, p < 0,001$.

Tiêu chí 2 (TC2): Hệ số tương quan gần như hoàn hảo, $\rho = 0,99, df = 24, p < 0,001$.

Tiêu chí 3 (TC3): Độ nhất quán tuyệt đối, $\rho = 1,00, df = 24, p < 0,001$.

Tiêu chí 4 (TC4): Hệ số tương quan rất cao, $\rho = 0,95, df = 24, p < 0,001$.

Tất cả các tiêu chí đều có hệ số tương quan **lớn hơn 0,90**, cho thấy mức độ nhất quán rất cao giữa hai lần chấm. Giá trị $p < 0,001$ trong tất cả các trường hợp khẳng định kết quả có ý nghĩa thống kê. Nhìn chung, kết quả này cho thấy phương pháp đánh giá có độ tin cậy test-retest rất cao, đảm bảo tính ổn định của dữ liệu thu thập (Cohen & cs, 2017).

4. Kết luận

4.1. Kết luận

Trong nghiên cứu này, chúng tôi đã dựa trên cơ sở các nghiên cứu trước để thiết kế một rubric đánh giá bài viết về năng lực giải quyết vấn đề toán học cho chủ đề “Giá trị phân số của một số” trong chương trình Toán 6. Thông qua việc thực hiện các đánh giá, kết quả cho thấy rubric này có thể áp dụng vào đánh giá trong thực tế để đo lường năng lực giải quyết vấn đề toán học của học sinh lớp 6 ở chủ đề “Giá trị phân số của một số”. Các tiêu chí và mức độ của rubric này cũng có thể sử dụng để đánh giá ở các nội dung khác trong chương trình Toán 6 mà có yêu cầu cần đạt về kỹ năng tương tự chủ đề “Phân số”.

4.2. Hạn chế và hướng phát triển

Do vấn đề thời gian thực nghiệm, nghiên cứu này có một số hạn chế nhất định. Thứ nhất, rubric này mới chỉ tập trung vào một chủ đề trong chương trình Toán 6. Thứ hai, mẫu thực nghiệm nhỏ và chỉ ở một trường, chưa bao quát các đối tượng học sinh khác nhau về các yếu tố nhân khẩu học như địa phương, dân tộc, nên việc đánh giá độ phân biệt chưa tổng quát cho các trường hợp khác.

Trong tương lai, chúng tôi sẽ dựa trên kết quả nghiên cứu này để điều chỉnh rubric, đồng thời tiến hành đánh giá một cách tổng quát hơn trong chương trình Toán 6 và ở một cỡ mẫu lớn hơn, để có thể đạt đến một thang đo tốt nhằm đánh giá các bài làm về năng lực giải quyết vấn đề toán học.

Tài liệu tham khảo

- Anisah, A., & Lastuti, S. (2018). Pengembangan Bahan Ajar berbasis HOTS untuk Meningkatkan Kemampuan Pemecahan Masalah Matematis Mahasiswa. *Kreano, Jurnal Matematika Kreatif-Inovatif*, 9(2), 191–197. <https://journal.unnes.ac.id/nju/kreano/article/view/16341>
- Asmana, A. T. (2018). PENGEMBANGAN RUBRIK ANALITIK UNTUK ASESMEN KOMUNIKASI MATEMATIKA TERTULIS DALAM PEMECAHAN MASALAH MATEMATIKA. *Jurnal Pembelajaran Matematika*, 5(1). <https://jurnal.uns.ac.id/jpm/article/view/26028>
- Bộ Giáo dục và Đào tạo. (2018). *Chương trình giáo dục phổ thông môn Toán*.
- Bryman, A. (2016). *Social research methods* (Fifth Edition). Oxford University Press.
- Bùi, A. K., Võ, T. T. P., Đỗ, T. P. T., & Bùi, P. T. H. (2023). *Giáo trình Đánh giá kết quả học tập Toán và Khoa học tự nhiên*. NXB Đại học Cần Thơ.
- Cahyadi, M. R., Maryanto, B. P. A., Syaifuddin, M., & Darmayanti, R. (2023). Development of Essay Test Assessment Rubric for Polya Theory-Based Mathematical Problem-Solving. *JNPM (Jurnal Nasional Pendidikan Matematika)*, 7(1), 167. <https://doi.org/10.33603/jnpm.v7i1.7724>
- Charles, R. I., Lester, F., & O'Daffer, P. G. (2008). *How to evaluate progress in problem solving* (8. print). The National Council of Teachers of Mathematics.
- Choirudin, C., Degeng, I. N. S., Kuswandi, D., Purnomo, P., & Maba, A. P. (2022). Career readiness among low-income Muslim students. *International Journal of Evaluation and Research in Education (IJERE)*, 11(3), 1400. <https://doi.org/10.11591/ijere.v11i3.22727>
- Cohen, L., Manion, L., & Morrison, K. (2017). *Research Methods in Education*. Routledge. <https://doi.org/10.4324/9781315456539>
- Darmayanti, R., Sugianto, R., & Muhammad, Y. (2022). Analysis of Students' Adaptive Reasoning Ability in Solving HOTS Problems Arithmetic Sequences and Series in Terms of Learning Style. *Numerical: Jurnal Matematika Dan Pendidikan Matematika*, 73–90. <https://doi.org/10.25217/numerical.v6i1.2340>
- Darmayanti, R., Syaifuddin, M., Rizki, N., Sugianto, R., & Hasanah, N. (2022). High school students' mathematical representation ability: Evaluation of disposition based on mastery learning assessment model (MLAM). *Journal of Advanced Sciences and Mathematics Education*, 2(1), 1–15. <https://doi.org/10.58524/jasme.v2i1.93>

- Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347–360. <https://doi.org/10.1080/02602938.2015.1111294>
- DeVellis, R. F. (2017). *Scale development: theory and applications* (Fourth edition). SAGE.
- Gunawardena, E. (2010). A rubric to self-assess and peer-assess mathematical problem solving tasks of college students. *Acta Didactica Napocensia*, 3(1), 75–88. <https://eric.ed.gov/?id=EJ1056109>
- Hoa, A. T., & Dương, P. V. (2024). Vận dụng mô hình 5E trong dạy học chủ đề “Đạo hàm cấp hai” (Toán 11) theo hướng phát triển năng lực giải quyết vấn đề toán học cho học sinh. *Tạp chí Giáo dục*, 24(20), 30–35. <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/2496>
- Hung, J. C., & Wang, C.-C. (2021). Exploring the website object layout of responsive web design: results of eye tracking evaluations. *The Journal of Supercomputing*, 77(1), 343–365. <https://doi.org/10.1007/s11227-020-03283-1>
- Jonan, Y. R. (2020). Pengembangan Rubrik Penskoran pada Asesmen Otentik untuk Materi Volume dan Luas Balok. *Journal of Medives : Journal of Mathematics Education IKIP Veteran Semarang*, 4(2), 275. <https://doi.org/10.31331/medivesveteran.v4i2.1174>
- Lamm, K., Lamm, A., & Edgar, D. (2020). Scale Development and Validation: Methodology and Recommendations. *Journal of International Agricultural and Extension Education*, 27(2), 24–35. <https://doi.org/10.5191/jiaee.2020.27224>
- Leacock, T., & Nesbit, J. (2007). A Framework for Evaluating the Quality of Multimedia Learning Resources. *Educational Technology and Society*, 10, 44–59. <https://www.sfu.ca/~jcn Nesbit/articles/LeacockNesbit2007.pdf>
- Nguyễn, H. H., Hoa, A. T., Lê, H. V., & Trịnh, T. L. M. (2023). Phát triển năng lực giải quyết vấn đề toán học cho học sinh lớp 10 khi dạy học chủ đề các hệ thức lượng trong tam giác. *Tạp chí Khoa học Trường Đại học Hồng Đức*, 63, 39–52. <https://vjol.info.vn/index.php/DHHD/article/view/87358>
- Nguyễn, T. T. T. (2025). Một số biện pháp phát triển năng lực tư duy và lập luận toán học cho học sinh trong dạy học chủ đề “Phân số” (Toán 6). *Tạp chí Giáo dục*, 25(2), 36–40. <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/2912>
- Nurhaifa, I., Hamdu, G., & Suryana, Y. (2020). Rubrik Penilaian Kinerja pada Pembelajaran STEM Berbasis Keterampilan 4C. *Indonesian Journal of Primary Education*, 4(1), 101–110. <https://doi.org/10.17509/ijpe.v4i1.24742>
- Nurtanto, M., Kholifah, N., Masek, A., Sudira, P., & Samsudin, A. (2021). Crucial problems in arranged the lesson plan of vocational teacher. *International Journal of Evaluation and Research in Education (IJERE)*, 10(1), 345. <https://doi.org/10.11591/ijere.v10i1.20604>
- Otgonbaatar, K. (2021). Effectiveness of anchoring vignettes in re-evaluating self-rated social and emotional skills in mathematics. *International Journal of Evaluation and Research in Education (IJERE)*, 10(1), 237. <https://doi.org/10.11591/ijere.v10i1.20716>

- Özer, B., Duran, V., & Tekke, M. (2020). Training of trainers: An action-based research for improving the pedagogical skills of academicians. *International Journal of Evaluation and Research in Education (IJERE)*, 9(3), 704. <https://doi.org/10.11591/ijere.v9i3.20327>
- Palinussa, A. L., Molle, J. S., & Gaspersz, M. (2021). Realistic mathematics education: Mathematical reasoning and communication skills in rural contexts. *International Journal of Evaluation and Research in Education (IJERE)*, 10(2), 522. <https://doi.org/10.11591/ijere.v10i2.20640>
- Phan, A. T. (2014). *Đánh giá năng lực giải quyết vấn đề của học sinh trong dạy học toán lớp 11 Trung học phổ thông* [Doctor Dissertation]. Trường Đại học Vinh.
- Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15(1), 18–39. <https://doi.org/10.1016/j.asw.2010.01.003>
- Salazar-Torres, J., Rincón Leal, O., & Vergel Ortega, M. (2021). The rubric as an assessment tool for solving problem situations in the physics and mathematics teaching context. *Journal of Physics: Conference Series*, 1981(1), 012018. <https://doi.org/10.1088/1742-6596/1981/1/012018>
- Suwarno, S., & Aeni, C. (2021). PENTINGNYA RUBRIK PENILAIAN DALAM PENGUKURAN KEJUJURAN PESERTA DIDIK. *Edukasi: Jurnal Pendidikan*, 19(1), 161. <https://doi.org/10.31571/edukasi.v19i1.2364>
- The Jamovi Project. (2024). *jamovi* (2.6). [Computer software] Retrieved from <https://www.jamovi.org>.
- Trần, T. Á. (2008). Một số nguyên tắc điều tra xã hội học. *Tạp chí Khoa học Trường Đại học Cần Thơ*, 9, 18–27. <https://ctujsvn.ctu.edu.vn/index.php/ctujsvn/article/view/533>
- Võ, X. M., & Tô, Q. L. (2023). Phát triển năng lực giải quyết vấn đề toán học cho học sinh lớp 10 qua dạy học giải bài tập phương trình quy về phương trình bậc hai. *Tạp chí Khoa học Trường Đại học Đồng Tháp*, 12(6), 3–11. <https://doi.org/10.52714/dthu.12.6.2023.1110>
- Wu, M. L. (2003). *The Application of Item Response Theory to Measure Problem-solving Proficiencies* [Doctor Dissertation]. The University of Melbourne.
- Zulfa, H., Saputro, D. R. S., & Riyadi. (2019). Students' difficulties in mathematics learning with artisan character type in HOTS trigonometry test. *Journal of Physics: Conference Series*, 1321(2), 022104. <https://doi.org/10.1088/1742-6596/1321/2/022104>