

TÍCH HỢP PHƯƠNG PHÁP ĐÁNH GIÁ SO SÁNH VÀO ĐÁNH GIÁ XÁC THỰC TRONG DẠY HỌC TOÁN

Lê Thúy Diệu* và Phạm Hà Mai Trang

Trường Đại học Sư phạm Hà Nội, Hà Nội, Việt Nam

*Tác giả liên hệ: Lê Thúy Diệu, e-mail: dieult@hnue.edu.vn

Ngày nhận bài: 8/1/2026. Ngày sửa bài: 24/2/2026. Ngày nhận đăng: 6/4/2026.

Tóm tắt. Việc triển khai một số thay đổi mới trong mô hình đánh giá khi phương pháp dạy học thay đổi để đáp ứng mục tiêu Chương trình Giáo dục phổ thông 2018 là điều tất yếu. Trong bối cảnh đó, đánh giá xác thực được xem là một cách tiếp cận phù hợp, chú trọng đánh giá năng lực vận dụng kiến thức toán học vào việc giải quyết các tình huống thực tiễn. Tuy nhiên, tính đa dạng của các nhiệm vụ xác thực đặt ra thách thức về tính nhất quán trong việc sử dụng thang đánh giá rubrics. Nghiên cứu này đề xuất quy trình gồm 6 bước sử dụng phương pháp đánh giá so sánh như một giải pháp kết hợp khả thi và tiến hành thử nghiệm sư phạm với đối tượng là sinh viên sư phạm trên một nhiệm vụ xác thực được thiết kế cho môn Toán lớp 11. Kết quả cho thấy đánh giá so sánh không chỉ đem lại độ tin cậy cao mà còn giúp học sinh phát triển năng lực đánh giá thông qua đánh giá đồng đẳng. Từ đó, nghiên cứu khuyến nghị tích hợp đánh giá so sánh vào hỗ trợ đánh giá xác thực để nâng cao hiệu quả dạy học môn Toán.

Từ khóa: đánh giá trong giáo dục, đánh giá so sánh, đánh giá xác thực, giáo dục toán học, độ tin cậy.

INTEGRATING COMPARATIVE JUDGMENT INTO AUTHENTIC ASSESSMENT IN MATHEMATICS EDUCATION

Le Thuy Dieu* and Pham Ha Mai Trang

Hanoi National University of Education, Hanoi, Vietnam

*Corresponding author: Le Thuy Dieu, e-mail: letd@upenn.edu

Received January 8, 2026. Revised February 24, 2026. Accepted April 6, 2026.

Abstract. Alongside renewed teaching methodology required by the 2018 Vietnamese General Education Curriculum, the implementation of a corresponding assessment model is essential. In this context, authentic assessment is considered an appropriate approach, as it focuses on evaluating students' ability to apply mathematical knowledge to real-world situations. However, the diversity of authentic tasks poses challenges to consistency when using fixed scoring rubrics. This study proposes a six-step procedure for applying comparative judgment as a feasible and integrative solution, and conducts a pedagogical trial with an authentic task in Grade 11 mathematics. The results indicate that comparative judgment not only ensures high reliability but also fosters students' evaluative competencies through peer assessment. Accordingly, the study recommends integrating comparative judgment into authentic assessment to enhance the effectiveness of mathematics teaching and learning.

Keywords: assessment in education, comparative judgment, authentic assessment, mathematics education, reliability.

1. Mở đầu

Chương trình Giáo dục phổ thông 2018 môn Toán đã nêu rõ: “Chương trình môn Toán chú trọng tính ứng dụng, gắn kết với thực tiễn hay các môn học, hoạt động giáo dục khác,... tạo cơ hội giúp học sinh vận dụng kiến thức, kỹ năng và kinh nghiệm của bản thân vào thực tiễn một cách sáng tạo”. Điều này đặt ra nhiệm vụ phải đổi mới phương pháp dạy học và song song với đó là một mô hình đánh giá phù hợp. Trong khi đó, các bài kiểm tra và kì thi truyền thống lại thường tập trung vào việc đánh giá khả năng sử dụng các thuật toán chuẩn để tìm ra câu trả lời đúng (Wilson, 1973). Thông thường, giáo viên thường đặt ra câu hỏi và đo lường câu trả lời của học sinh bằng cách so sánh với câu trả lời hoặc kết quả lí tưởng (thường gọi là đáp án) hoặc bộ tiêu chí xác định mức độ kỹ lưỡng/chính xác mà người học đã trả lời các câu hỏi. Dù có ưu điểm là dễ chấm điểm và đảm bảo độ tin cậy cao trong đánh giá (McConlogue, 2012), công cụ này lại có thể khiến học sinh trở thành người học thụ động, chú trọng ghi nhớ các thuật toán hơn là hiểu bản chất toán học (Flores & nnk, 2015). Để khắc phục những hạn chế này, Svinicki (2004) đề xuất rằng cần kết hợp các nhiệm vụ đánh giá xác thực vào quá trình đánh giá học sinh. Tuy nhiên, một vấn đề quan trọng nảy sinh khi áp dụng phương pháp đánh giá xác thực là độ tin cậy khi đánh giá các nhiệm vụ xác thực. “Các sản phẩm khác nhau từ các học sinh khác nhau có thể được chấm điểm một cách nhất quán và có thể so sánh không?”. Việc sử dụng một tiêu chí đánh giá nêu rõ các tiêu chí chấm điểm như hiện nay chúng ta đang làm khó có thể áp dụng đối với một nhiệm vụ mở như vậy, một phần là do sự đa dạng các cách tiếp cận khác nhau mà học sinh có thể áp dụng để tiếp cận nhiệm vụ (Cantley, 2023). Để giải quyết vấn đề này, Ian Cantley (2023) đã đề xuất phương pháp “đánh giá so sánh” của Pollitt như một biện pháp khắc phục. Đánh giá so sánh đã được sử dụng trong đánh giá tổng kết với các hình thức kiểm tra viết tiếng Anh, hiểu biết khái niệm trong toán học,... và đã được chứng minh độ tin cậy cao trong nhiều nghiên cứu (Bartholomew & Yoshikawa, 2018; Jones & Inglis, 2015; Pollitt, 2012). Bên cạnh đó, đánh giá so sánh được tiếp cận không chỉ như một giải pháp kỹ thuật nhằm nâng cao độ tin cậy của đánh giá xác thực, mà còn như một hoạt động có tiềm năng hỗ trợ quá trình học tập của người học. Những người học tham gia vào quá trình đánh giá so sánh đều cho rằng họ nhận được những phản hồi hữu ích hơn so với các cách thức phản hồi khác mà sinh viên đã trải nghiệm (Ellison, 2023). Tuy nhiên, các nghiên cứu trên chủ yếu được đặt trong bối cảnh đánh giá tổng kết. Nghiên cứu của Van Daal & nnk (2019), Pollitt & nnk (2012) chỉ đặt trong bối cảnh đánh giá bài kiểm tra viết, nghiên cứu của Bisson đánh giá hiểu biết toán học của học sinh (2016). Xét trong hoạt động đánh giá xác thực môn Toán, Ian Cantley mới dừng lại ở việc đề xuất giải pháp mà chưa có minh họa cụ thể, vì vậy, nghiên cứu này có thể coi là tiếp nối hướng nghiên cứu của ông nhằm đưa ra quy trình cụ thể để áp dụng và kiểm chứng hiệu quả của đánh giá so sánh đối với đánh giá xác thực trong dạy học môn Toán.

2. Cơ sở lí luận

2.1. Đánh giá xác thực trong dạy học môn Toán

Theo Mueller (2005), đánh giá xác thực là một hình thức đánh giá trong đó học sinh được yêu cầu thực hiện các nhiệm vụ trong thế giới thực để chứng tỏ việc áp dụng có ý nghĩa các kiến thức và kỹ năng cần thiết. Như vậy, có thể xem đánh giá xác thực trong dạy học môn toán là quá trình đánh giá khả năng áp dụng linh hoạt và sáng tạo các kiến thức và kỹ năng toán học để giải quyết các vấn đề trong thực tiễn, trình bày các giải pháp dưới dạng biểu diễn toán học. Qua đó thể hiện sự hiểu biết sâu sắc và khả năng áp dụng toán học một cách hiệu quả và có ý nghĩa. Dựa trên sự tổng hợp các nghiên cứu của các tác giả Wiggins (1990), Mueller (2005), Svinicki (2004), đánh giá xác thực có các đặc trưng sau:

Thứ nhất, đánh giá xác thực yêu cầu học sinh phải tạo ra sản phẩm hoặc sự thể hiện có chất lượng qua việc phối hợp các kiến thức, kỹ năng tư duy bậc cao để giải quyết vấn đề.

Thứ hai, đánh giá được thực hiện trong những tình huống càng giống với bối cảnh thực hiện các kỹ năng liên quan càng tốt.

Thứ ba, nội dung và yêu cầu của nhiệm vụ xác thực phải gắn với thực tế, chứa đựng những vấn đề thực tế mà HS có thể gặp phải sau này.

Thứ tư, đánh giá đòi hỏi sự phán đoán và sáng tạo, tạo ra những điều mới mẻ trong quá trình thực hiện, dựa trên việc giải quyết các vấn đề không có cấu trúc mà có thể có nhiều hơn một câu trả lời đúng.

Thứ năm, học sinh là chủ thể của quá trình thực hiện nhiệm vụ, cần có sự phối hợp, tương tác xã hội, sự phụ thuộc tích cực của các thành viên và trách nhiệm cá nhân của từng thành viên. Học sinh được tự đánh giá, theo dõi sự tiến bộ hoặc tiến độ học tập của cá nhân.

Thứ sáu, cấu trúc sản phẩm của nhiệm vụ xác thực không chỉ được tạo ra bởi giáo viên mà còn được tạo ra bởi học sinh. Giáo viên là người thiết kế, nhưng học sinh cũng có thể lựa chọn, xác định được cái gì sẽ được sử dụng, tạo ra và thể hiện như một minh chứng cho sự thành thạo.

2.2. Vấn đề về độ tin cậy trong đánh giá xác thực

Mặc dù các nhiệm vụ mở phát huy sức sáng tạo và khả năng giải quyết vấn đề của học sinh, nhưng cũng tạo ra sự khó khăn nhất định trong việc chấm điểm một cách nhất quán, tức là ảnh hưởng đến độ tin cậy. “Độ tin cậy là tính ổn định của kết quả đánh giá, tức là có điểm số tương tự giữa các lần đo khi không có những tác động đáng kể” (Nguyễn & Đào, 2015, tr. 82).

Đối với những nhiệm vụ xác thực, học sinh sẽ thực hiện một công việc để giải quyết vấn đề thay vì đưa ra một câu trả lời đúng. Vì vậy, có thể có nhiều hơn một câu trả lời được chấp nhận, điều này dẫn đến sự đa dạng và không thể đoán trước của các câu trả lời mà học sinh đưa ra (Svinicki, 2004). Với đánh giá dựa theo tiêu chí, một phiếu hướng dẫn đánh giá được thiết kế tốt phải trả lời được câu hỏi “Các thông tin mô tả có rõ ràng không?” (Nguyễn & Đào, 2015, tr. 50). Trong khi đó, việc sử dụng một tiêu chí đánh giá nêu rõ các tiêu chí chấm điểm khó có thể áp dụng đối với một nhiệm vụ mở như vậy. Hơn nữa, việc làm cho tiêu chí cụ thể hơn có thể hạn chế không gian cho sự sáng tạo và tư duy phản biện trong việc tiếp cận nhiệm vụ (Rezaei & Lovorn, 2010).

Bloxham & Price (2015) đã chỉ ra rằng việc chấm điểm các nhiệm vụ khác nhau mang tính chủ quan cao, chủ yếu do sự khác biệt về mức độ khắt khe hoặc dễ tính. Kinh nghiệm, sở thích cá nhân và nhiều yếu tố khác đều ảnh hưởng đến kết quả đánh giá và góp phần gây khó khăn trong việc chấm điểm một cách đáng tin cậy giữa những người đánh giá. Khi một giám khảo chấm điểm cho một sản phẩm bằng tiêu chí đánh giá, họ cần phải so sánh nó với những sản phẩm tương tự trong đầu, mở đường cho ba nguồn lỗi có thể xảy ra: sự khác biệt trong tiêu chuẩn cá nhân của họ, trong cách họ phân biệt chính xác các phẩm chất mà họ đang mong đợi và trong cách diễn giải riêng của họ về tiêu chí (Pollitt & Whitehouse, 2012). Do đó, cách tiếp cận này khiến học sinh gặp khó khăn trong việc hình thành cái nhìn toàn diện về chất lượng của sản phẩm. Hơn nữa, nó hạn chế học sinh phát triển các kỹ năng tự xác định các tiêu chí phù hợp với một nhiệm vụ nhất định. Điều quan trọng là các kỹ năng này cần được phát triển ở học sinh để các em sẵn sàng cho cuộc sống bên ngoài trường học, nơi không có tiêu chí nào được xác định trước (van Daal & nnk, 2023).

3. Tiềm năng của đánh giá so sánh trong việc làm tăng độ tin cậy của đánh giá xác thực

3.1. Đánh giá so sánh và tiềm năng của đánh giá so sánh trong việc làm tăng độ tin cậy của đánh giá xác thực

Để khắc phục khó khăn trong việc đảm bảo độ tin cậy khi đánh giá các nhiệm vụ mở, Cantley (2023) đã cân nhắc đến một phương pháp đánh giá được gọi là đánh giá so sánh (CJ - Comparative Judgment). Cơ chế của đánh giá so sánh rất đơn giản: những người chấm điểm đặt hai đối tượng để so sánh với nhau và quyết định đối tượng nào thể hiện tốt hơn trên một số tiêu chí hoặc cấu trúc toàn bộ, lần lượt đến hết. Sau đó, các quyết định được đưa vào một mô hình thống kê (ví dụ: Bradley-Terry-Luce) để đưa ra ước tính tham số chuẩn hóa cho mỗi học sinh và xây dựng thứ hạng.

Nhà tâm lý học thực nghiệm Laming (2003) tuyên bố: "Không có đánh giá tuyệt đối. Tất cả các đánh giá đều là so sánh một thứ này với một thứ khác". Cơ sở tâm lý này cho rằng con người rất nhất quán khi so sánh một vật thể với vật thể khác, ngay cả khi tính chất được so sánh không

thể được xác định hoặc đo lường một cách khách quan. Ví dụ, rất khó để đánh giá chính xác một vật nặng bao nhiêu gam bằng cách cầm nó trên một tay. Ngược lại, tương đối dễ để đánh giá vật nào trong hai vật nặng hơn bằng cách cầm một vật ở mỗi tay.

Độ tin cậy của CJ thường được xác định bởi Độ tin cậy tách thang đo (SSR - Scale Separation Reliability). SSR có thể được coi là thước đo tính nhất quán của những người đánh giá. Độ tin cậy giữa các giám khảo càng thấp thì kết quả của một học sinh cụ thể càng phụ thuộc vào tính cách riêng của bất kỳ ai tình cờ chấm bài làm, và do đó, đánh giá càng kém công bằng (Jones & Inglis, 2015).

Jones & Inglis (2015) đã chỉ ra rằng việc sử dụng đánh giá so sánh trong toán học có ý nghĩa đối với cách thiết kế các câu hỏi kiểm tra. Họ thấy rằng khi không cần quan tâm đến việc chấm điểm, những người ra đề thi đã tạo ra một bài kiểm tra với nhiều câu hỏi mở hơn, ít có cấu trúc hơn so với các bài kiểm tra ở trường hiện nay. Ngoài ra, Jones & Inglis (2015) cũng kết luận quy trình đánh giá so sánh mang lại kết quả đánh giá có độ tin cậy cao giữa những người đánh giá.

Bartholomew (2018) đã tổng hợp các nghiên cứu xung quanh đánh giá so sánh cũng tin rằng với độ tin cậy cao hơn đáng kể so với các hình thức đánh giá khác, cùng với các lợi ích khác trong đánh giá đồng đẳng, đánh giá so sánh là một công cụ mạnh mẽ có khả năng cách mạng hóa việc đánh giá vấn đề mở trong các bối cảnh giáo dục phổ thông.

3.2. Thực trạng mức độ nhận diện và phổ biến của đánh giá so sánh trong giáo dục

Nhằm phân tích bối cảnh ban đầu về mức độ nhận diện và mức độ phổ biến của đánh giá so sánh trong thực tiễn giáo dục, nhóm tác giả đã tiến hành một khảo sát thăm dò đối với các cá nhân có liên hệ trực tiếp hoặc gián tiếp với hoạt động dạy và học. Khảo sát được thực hiện thông qua bảng hỏi trực tuyến và phát tán theo phương pháp chọn mẫu thuận tiện. Bảng hỏi gồm các câu hỏi chính sau: (1) Bạn có phải là giáo viên/công tác trong ngành giáo dục không? (bao gồm cả sinh viên ngành sư phạm); (2) Bạn đã từng biết về đánh giá so sánh chưa? (đã có mô tả sơ lược về đánh giá so sánh); (3) Trong các lớp học của bạn, giáo viên đã từng sử dụng đánh giá so sánh chưa? (Nếu là giáo viên, thì bạn đã từng sử dụng đánh giá so sánh với lớp học của mình chưa?); (4) Bạn nghĩ đây là phương pháp hiệu quả để đánh giá trong lớp học? (5) Ngoài lớp học, bạn đã gặp/sử dụng loại đánh giá so sánh này chưa? Sử dụng ở mức độ như thế nào?

Chúng tôi tiến hành khảo sát trên 67 người, trong đó có 50,7% là giáo viên, sinh viên sư phạm hoặc người đang công tác trong lĩnh vực giáo dục; số còn lại là những người không làm việc trực tiếp trong ngành giáo dục nhưng đã từng trải qua quá trình học tập trong hệ thống giáo dục phổ thông và đại học tại Việt Nam. Cơ cấu mẫu này cho phép khảo sát tiếp cận các góc nhìn đa dạng từ cả phía người dạy và người học, qua đó cung cấp bức tranh tổng quan về mức độ nhận diện của đánh giá so sánh trong nhận thức và trải nghiệm giáo dục. Dữ liệu cho thấy đánh giá so sánh tuy không hoàn toàn xa lạ, song vẫn chưa được khai thác đúng mức trong quá trình dạy và học. Cụ thể, có đến 79,1% người tham gia cho biết đã từng biết đến đánh giá so sánh, phản ánh một mức độ nhận diện tương đối tốt. Và có 68,7% người tham gia cho biết rằng trong lớp học của họ đã từng sử dụng phương pháp này. Bên cạnh đó, kết quả khảo sát cho thấy hơn một nửa số người tham gia (59,7%) đánh giá đây là phương pháp hiệu quả/phù hợp để áp dụng trong lớp học, trong khi có 20 người (29,9%) không chắc chắn với nhận định này. Con số này cũng có thể phần nào khẳng định tiềm năng của phương pháp này nếu được triển khai đúng cách. Đáng chú ý, trong đời sống ngoài lớp học, có tới 91,1% người đã từng gặp hoặc sử dụng hình thức đánh giá so sánh (bao gồm các mức độ nhiều và ít), cho thấy đây là một phương pháp đánh giá có tính gần gũi, dễ tiếp cận và đã xuất hiện trong các hoạt động thường nhật của người học. Tuy nhiên, hình thức này chưa thực sự được nhận diện rõ dưới tên gọi học thuật trong giáo dục tại Việt Nam.

Từ thực trạng nêu trên, có thể thấy người học và giáo viên đã có sự tiếp xúc nhất định với hình thức đánh giá so sánh ở các mức độ khác nhau, điều này cho thấy tính khả thi cao nếu có một quy trình phù hợp được đề xuất và hỗ trợ triển khai.

4. Sử dụng kết hợp đánh giá đồng đẳng trong đánh giá so sánh và bàn luận

Nhiều nghiên cứu đã chỉ ra rằng đánh giá so sánh có thể được sử dụng như một công cụ đánh giá đồng đẳng có tính giá trị và đáng tin cậy (Jones & Alcock, 2014; Jones & Wheadon, 2015; Seery & nnk, 2012), nghĩa là so sánh có thể được thực hiện bởi chính học sinh.

Bên cạnh phản hồi từ giáo viên, nghiên cứu cũng khai thác phản hồi từ học sinh – một dạng học tập qua đánh giá ngày càng được nhấn mạnh trong lý thuyết giáo dục hiện đại. Trong nghiên cứu này, đánh giá so sánh còn được tiếp cận ở góc nhìn theo triết lý “đánh giá là học tập” (Assessment as Learning), đánh giá không chỉ nhằm xếp loại người học mà còn là một hoạt động của quá trình học tập. “Mục đích đánh giá này tập trung vào vai trò kết nối quan trọng của người học giữa đánh giá và học tập, xác định vai trò tích cực và chủ động của người học trong quá trình đánh giá, liên hệ các kiến thức đã có và dùng những kiến thức đó để học hỏi các kiến thức mới thì đây chính là quá trình siêu nhận thức” (Đặng, 2022, tr. 8). Trong bối cảnh đó, đánh giá so sánh vừa là phương thức đánh giá đáng tin cậy, vừa là trải nghiệm học tập tích cực cho học sinh.

Một nhược điểm của phương pháp đánh giá so sánh do giáo viên làm giám khảo là học sinh không nhận được phản hồi chi tiết về bài tập đã nộp (Jones & Alcock, 2014; Jones & Wheadon, 2015). “Tuy nhiên, mặc dù không có phản hồi nào được trình bày theo nghĩa truyền thống, chúng tôi lập luận rằng quá trình đánh giá thu hút người học vào các so sánh có ý nghĩa về chất lượng câu trả lời và do đó cung cấp một hình thức phản hồi mới lạ và có lợi về thành tích của chính họ” (Jones & Sirl, 2017, tr.160). Thông qua việc đánh giá hoạt động, sản phẩm học tập của bạn bè, học sinh có thể học hỏi những điểm tốt hoặc rút kinh nghiệm từ những điểm chưa tốt của bạn, hình thành rõ ràng hơn trong bản thân mình các yêu cầu của nhiệm vụ học tập.

Đánh giá so sánh yêu cầu học sinh đưa ra đánh giá toàn diện, tức là học sinh đánh giá các sản phẩm như một tổng thể và trực tiếp đưa ra phán đoán chung (Pollitt, 2012). Ngoài ra, đánh giá so sánh mang đến cho học sinh cơ hội suy ngẫm về cách khái niệm hóa chất lượng của một tác phẩm (Sadler, 2009). Nhiều nghiên cứu đã chỉ ra rằng kết quả học tập cao hơn khi so sánh các ví dụ so với việc xem xét các ví dụ riêng lẻ. Trong khi so sánh, học sinh tìm kiếm điểm tương đồng và khác biệt giữa hai tác phẩm, làm nổi bật các khía cạnh khác nhau của mỗi tác phẩm. Theo cách này, học sinh sẽ hiểu rõ hơn về các đặc điểm quan trọng mà một tác phẩm tốt phải đáp ứng (Alfieri và nnk, 2013; Pachur & Olsson, 2012). Nghiên cứu của Bouwer (2018) cũng đã làm sáng tỏ cách phân tích các ví dụ có chất lượng khác nhau có thể thúc đẩy sự hiểu biết và chất lượng bài làm của học sinh. Ngoài ra, học sinh đã thể hiện sự hứng thú trong việc phân tích bài làm của bạn, bày tỏ thích được học hỏi từ việc đối chiếu, và chủ động đề xuất tiêu chí đánh giá rõ ràng hơn. Đây là dấu hiệu cho thấy đánh giá so sánh đang hoạt động như một hình thức học tập hiệu quả.

Sau khi hoàn thành đánh giá, thay vì chỉ công bố điểm số hoặc vị trí xếp hạng, giáo viên có thể chia sẻ các bài làm mẫu thuộc nhiều mức chất lượng khác nhau để học sinh thảo luận. Việc đặt các bài “tốt hơn” và “chưa tốt bằng” cạnh nhau, kèm theo bình luận cụ thể giúp học sinh hiểu rõ tiêu chí đánh giá và cách để cải thiện bài làm của chính mình. Ví dụ: bài A được đánh giá cao hơn bài B vì trình bày trực quan hơn, giải pháp khả thi hơn,... Quá trình này làm cho học sinh vừa là người đưa ra phản hồi, vừa là người nhận thông tin, phản biện và điều chỉnh.

Tuy nhiên, các giám khảo, tức là những học sinh tham gia thực hiện đánh giá so sánh trong quá trình đánh giá đồng đẳng cần phải thống nhất về các đặc điểm của một sản phẩm tốt hơn sản phẩm khác để đảm bảo độ tin cậy và tính giá trị. Chính vì vậy, *đào tạo giám khảo* là việc cần phải làm trước khi bắt đầu tiến hành đánh giá, nhằm đảm bảo họ có sự hiểu biết chung và đầy đủ về mục tiêu đánh giá, tiêu chí chất lượng đối với nhiệm vụ học tập. Trong bối cảnh học sinh tham gia vào quá trình đánh giá, việc đào tạo này càng trở nên quan trọng, vì năng lực đánh giá của người học có thể không đồng đều, dẫn đến kết quả có thể bị ảnh hưởng.

4.1. Công cụ No More Marking và quy trình áp dụng đánh giá so sánh vào đánh giá xác thực

No More Marking là một nền tảng trực tuyến hoạt động trong việc xuất bản nghiên cứu học thuật về đánh giá so sánh và cung cấp giải pháp đánh giá bài viết dựa trên phương pháp này. Công cụ này cho phép nhiều người cùng tham gia đánh giá, các phản hồi họ đưa ra sau đó được hợp nhất thành thứ hạng của tất cả các sản phẩm. Thứ hạng bao gồm cả vị trí của nó và sự khác biệt tương đối về chất lượng giữa các sản phẩm. (<https://www.nomoremarking.com/>)

Qua những phân tích trên, tác giả đề xuất một quy trình đầy đủ để áp dụng đánh giá so sánh vào đánh giá xác thực gồm 6 bước:

Bước 1: Thiết kế nhiệm vụ đánh giá xác thực.

Bước 2: Thảo luận giữa các giám khảo về các tiêu chí đánh giá đối với nhiệm vụ học tập (trong trường hợp này là giáo viên và học sinh).

Giáo viên tạo điều kiện để học sinh tham gia thảo luận và thống nhất các tiêu chí đánh giá cốt lõi của nhiệm vụ học tập. Hoạt động này nhằm hình thành nhận thức chung về chất lượng của sản phẩm học tập. Trong giai đoạn này, các tiêu chí được xây dựng ở mức định hướng, đóng vai trò như một khung tham chiếu chung, chưa nhất thiết được chuẩn hóa hoàn toàn hoặc sử dụng trực tiếp cho việc ra quyết định đánh giá. Việc hiệu chỉnh, làm rõ cách diễn giải và thống nhất cách vận dụng các tiêu chí sẽ được thực hiện ở bước đào tạo giám khảo (Bước 4).

Bước 3: Tổ chức cho học sinh thực hiện nhiệm vụ học tập.

Giáo viên tổ chức cho học sinh thực hiện nhiệm vụ theo nhóm hoặc cá nhân. Trong quá trình này, học sinh sẽ tìm hiểu tài liệu, thu thập thông tin; thực hành giải quyết vấn đề, tạo ra sản phẩm học tập. Giáo viên đóng vai trò là người hỗ trợ, không can thiệp quá sâu.

Bước 4: Đào tạo giám khảo trước khi tiến hành đánh giá so sánh

Bước này tập trung vào thống nhất tiêu chí ở mức độ chi tiết hơn và hướng dẫn thao tác thực hành: giám khảo được hướng dẫn cách ra quyết định so sánh trên nền tảng (ví dụ: No More Marking), thực hành trên một tập bài mẫu, thảo luận các trường hợp biên và thống nhất cách diễn giải các khía cạnh chất lượng trước khi bắt đầu chấm chính thức.

Bước 5: Tiến hành đánh giá đồng đẳng bằng phương thức đánh giá so sánh

Bước 6: Phản hồi sau đánh giá



Hình 1. Quy trình áp dụng đánh giá so sánh vào đánh giá xác thực

4.2. Minh họa việc sử dụng đánh giá so sánh với một nhiệm vụ xác thực cụ thể môn Toán

Bước 1. Thiết kế nhiệm vụ đánh giá xác thực.

Tác giả tiến hành xây dựng một nhiệm vụ xác thực ở cấp THPT – “Một số hình thức tích lũy và đầu tư tài chính”. Nhiệm vụ được thực hiện theo nhóm trong thời gian 2 tuần và báo cáo sản phẩm bằng hình thức thuyết trình.

Nội dung: Với mức thu nhập 30 triệu/tháng (sau khi trừ thuế thu nhập cá nhân), anh An dành một phần thu nhập để trang trải chi phí sinh hoạt (10 triệu/tháng) và phần tiền còn lại hoặc đầu tư vào các hình thức tài chính khác nhau nhằm tích lũy vốn để mua một căn hộ sau 5 năm (khoảng 2 tỉ). Hiện tại anh đã có một khoản tiền 320 triệu. Hãy lập kế hoạch tiết kiệm và đầu tư phù hợp để anh An thực hiện được mục tiêu.

Đối tượng thực hiện nhiệm vụ trong thử nghiệm sư phạm minh họa: 7 nhóm sinh viên sư phạm (năm thứ ba) trong học phần Tổ chức dạy học môn Toán (6 sinh viên/nhóm) ở Trường Đại học Sư phạm Hà Nội. Việc lựa chọn sinh viên nhằm kiểm tra tính khả thi của quy trình và cách

thức vận hành của đánh giá so sánh trong bối cảnh một nhiệm vụ xác thực. Nếu phương pháp này hoạt động hiệu quả đối với các sản phẩm của sinh viên, nó cũng sẽ mang lại kết quả tương tự đối với các sản phẩm của học sinh.

Mô tả chi tiết nhiệm vụ:

- Tìm hiểu thông tin về một số cổ phiếu, trái phiếu của doanh nghiệp niêm yết trên thị trường chứng khoán Việt Nam qua hai trang web <https://www.hnx.vn/> và <https://www.hsx.vn/> hoặc những trang web của công ty chứng khoán; lãi suất gửi tiết kiệm của một số ngân hàng.

- Phân tích ưu điểm, nhược điểm của từng hình thức (gửi tiết kiệm tích lũy, đầu tư chứng khoán thông qua cổ phiếu, trái phiếu). Đánh giá tiềm năng tăng trưởng và rủi ro của 3-4 cổ phiếu, trái phiếu dựa vào một số chỉ số như: vốn hóa, EPS, P/E,...

- Chọn hình thức tích lũy hoặc đầu tư phù hợp, phân bổ nguồn vốn hợp lý để đạt mục tiêu. Trình bày kế hoạch đầu tư rõ ràng (số tiền đầu tư ban đầu, kỳ hạn, ước tính lợi nhuận, sử dụng công cụ tính toán phù hợp - có giải thích)

- Đề xuất giải pháp điều chỉnh kế hoạch khi cần thiết (Ví dụ: thị trường chứng khoán không thuận lợi, giá nhà tăng vọt,...)

Bước 2. Thảo luận giữa các giám khảo về các tiêu chí đánh giá đối với nhiệm vụ học tập.

Ở bước này, giáo viên và học sinh cùng thảo luận và thống nhất các tiêu chí cốt lõi phản ánh chất lượng của sản phẩm học tập. Kết quả của bước thảo luận là sự thống nhất tương đối về một số khía cạnh cốt lõi cần được xem xét khi đánh giá sản phẩm, bao gồm: (i) mức độ đầy đủ và độ tin cậy của thông tin thu thập; (ii) tính hợp lý trong việc lựa chọn và phân bổ nguồn lực; (iii) mức độ rõ ràng và khả thi của kế hoạch đề xuất; (iv) cách thức sử dụng các công cụ hỗ trợ và lập luận kèm theo; (v) khả năng điều chỉnh và cải tiến kế hoạch để đạt mục tiêu.

Bước 3. Tổ chức cho các nhóm sinh viên thực hiện nhiệm vụ học tập.

Bước 4. Đào tạo giám khảo trước khi tiến hành đánh giá so sánh.

Trên cơ sở các tiêu chí đã được thống nhất ở Bước 2, giám khảo được đào tạo và hiệu chuẩn thông qua việc sử dụng một phiếu hướng dẫn đánh giá theo tiêu chí. Bảng 1 trình bày phiếu hướng dẫn đánh giá được sử dụng trong giai đoạn đào tạo giám khảo. Phiếu hướng dẫn đánh giá theo tiêu chí này đóng vai trò hỗ trợ giám khảo trong việc nhận diện, thảo luận và hiệu chuẩn cách hiểu về chất lượng tổng thể của sản phẩm khi thực hiện đánh giá so sánh.

Bảng 1. Phiếu hướng dẫn đánh giá theo tiêu chí sử dụng trong đào tạo giám khảo

Tiêu chí	1	2	3
<i>Tìm hiểu thông tin cổ phiếu</i>	Tìm hiểu thông tin sơ sài, số lượng ít, có thể có thông tin sai lệch, không cập nhật	Tìm hiểu khá đầy đủ về 3 - 4 cổ phiếu, chưa đa dạng nhóm ngành, chưa diễn giải xác đáng về các chỉ số	Tìm hiểu kỹ về đa dạng các mã cổ phiếu, trái phiếu, thông tin đầy đủ, chính xác, có diễn giải các chỉ số, có cập nhật theo thời gian
<i>Lựa chọn và phân bổ vốn</i>	Lựa chọn và phân bổ vốn chưa hợp lý, không đạt mục tiêu, không giải thích.	Lựa chọn và phân bổ vốn tương đối hợp lý, đạt mục tiêu nhưng chưa giải thích rõ.	Lựa chọn kênh đầu tư và phân bổ vốn hợp lý, tối ưu, đạt mục tiêu, đưa ra giải thích rõ ràng.
<i>Xây dựng kế hoạch chi tiết</i>	Kế hoạch sơ sài, thiếu nhiều yếu tố quan trọng	Kế hoạch khá chi tiết, nhưng còn thiếu một vài yếu tố	Kế hoạch chi tiết, đầy đủ, tường minh các yếu tố
<i>Công cụ tính toán (công thức, hàm tính toán)</i>	Chưa có công cụ tính toán, các con số chưa có cơ sở	Có sử dụng được công cụ tính toán, nhưng còn sai sót	Sử dụng chính xác các công cụ tính toán, lí giải xác đáng lí do sử dụng công cụ
<i>Điều chỉnh kế hoạch</i>	Không đề xuất điều chỉnh/ đề xuất không phù hợp.	Đề xuất điều chỉnh, nhưng giải thích chưa rõ ràng.	Đề xuất các điều chỉnh hợp lí để đạt mục tiêu, giải thích rõ ràng.

Bước 5: Tiến hành đánh giá đồng đẳng bằng phương thức đánh giá so sánh

Sử dụng công cụ No More Marking trực tuyến. Khi hoàn thành việc đánh giá, trang web sẽ trả về kết quả và độ tin cậy chung của nhiệm vụ và bảng tính kết quả như hình dưới đây:

Candidate	QRCode	FirstName	Last Name	Gender	Group	Mark	Scaled Score	Scaled Score SE	True Score	True Score SE	Infit	Score	Prop Score	Local Comparisons
* SIFTMH	NHOM	1	M	11	*	38	2.581068197147076	0.047603772	0.917511709111	0.7258092817636406	1.9727272727272727	5.442148760330577		22
* UGG9DR	NHOM	2	M	11	*	78	4.112070041789381	1.954356305789025	5.087238293131	1.0316944474689296	3.8799999999999999	0.694		20
* MAAUXX	NHOM	4	M	11	*	6	1.182452391745961	-1.42476081	2.367752833311	1.130006378029472	7.0899999999999999	35.44999999999999		20
* OGGQGG	NHOM	5	F	11	*	0	1.139809083714502	-1.737088769	3377594003611	0.971534546198158	1.157142857142857	0.245578231		21
* PPFXF9	NHOM	7	F	11	*	7	1.220808942963679	-1.382816202	7170841191341	0.7568152944011503	5.157142857142857	0.245578231		21
* FAID9T	NHOM	8	M	11	*	28	2.186875340802541	-0.403234885	7071494990551	1.1069556390613233	5.9699999999999999	0.5485		20
* BCBUEE	NHOM	0	F	11	*	100	16.312042479061752	3.945940587885666	6.389773801372	0.8420300423744468	17.76	1.8880000000000000		20

Hình 2. Kết quả tải xuống ở định dạng csv

Bảng kết quả bao gồm một số thông tin mà ta cần quan tâm như:

- **Scaled Score:** Thang điểm đã điều chỉnh cho sản phẩm, điểm càng cao thì chất lượng tổng thể (theo đánh giá so sánh) càng tốt. Giới hạn thang điểm được xác định trong phần cài đặt từ đầu.
- **True Score:** Điểm thực tế được sử dụng bởi các phép tính thống kê liên quan
- **Infit:** phản ánh mức độ không đồng tình giữa các giám khảo về chất lượng của sản phẩm. Giá trị 1,3 trở lên có nghĩa là có sự bất đồng đáng kể liên quan đến tập lệnh của học sinh này.
- **Local Comparisons:** Số lượng so sánh cục bộ đã được thực hiện cho tập lệnh này.

Nhiệm vụ này có độ tin cậy là 0.87 với 8 giám khảo (bao gồm 7 nhóm sinh viên và 1 chuyên gia) và 20 so sánh trên mỗi sản phẩm. Giá trị này là đủ tốt, cho thấy phương pháp này có tiềm năng để áp dụng vào quá trình đánh giá trong dạy học môn Toán.

Bước 6: Phản hồi sau đánh giá.

Sau khi đánh giá, các nhóm sinh viên sẽ thảo luận bài làm mẫu của các nhóm 0, 5, 8 để cùng thảo luận vì sao sản phẩm 0 được đánh giá tốt nhất, sản phẩm 5 được đánh giá chưa tốt bằng các sản phẩm còn lại,... và khó khăn trong quá trình đánh giá.

Bên cạnh dữ liệu định lượng thu được từ kết quả đánh giá so sánh trên nền tảng *No More Marking*, nghiên cứu còn sử dụng các nguồn dữ liệu định tính nhằm hỗ trợ diễn giải kết quả và làm rõ tác động học tập của quá trình đánh giá. Cụ thể, dữ liệu định tính được thu thập thông qua (i) quan sát quá trình sinh viên tham gia đánh giá đồng đẳng và thảo luận sau đánh giá; (ii) ý kiến phản hồi của sinh viên được ghi nhận trong các buổi thảo luận ở Bước 6; và (iii) các nhận xét tự phản ánh của sinh viên về nhiệm vụ và quá trình đánh giá. Các dữ liệu này được sử dụng để nhận diện những thay đổi trong cách sinh viên diễn đạt yêu cầu của nhiệm vụ và cách họ lí giải chất lượng của sản phẩm học tập.

Một số câu trả lời đã nhận được: Thứ nhất, về các đặc điểm của một sản phẩm mà họ đánh giá tốt hơn nhưng sản phẩm khác: “Sản phẩm có sự so sánh đa dạng các lựa chọn, cho phép nhà đầu tư có góc nhìn tổng quát”; “Đúng trọng tâm, đầy đủ, bố cục rõ ràng, nhưng chủ yếu vẫn là lĩnh vực đầu tư cũng như rủi ro dựa trên đầy đủ thông tin hay không còn những cái khác thì không quá quan trọng”; “Trình bày đẹp, bố cục rõ, có bảng số liệu thống kê chi tiết. Số liệu bên trong được nhìn nhận tương quan từ những ứng dụng đầu tư”. Thứ hai, về khó khăn trong quá trình đánh giá: “Chưa quen với giao diện nên quá trình đánh giá tốn nhiều thời gian”; “Vì phần nhìn trên web hạn chế, nên có một số nhóm có thể bị chọn nhầm hoặc bỏ sót, dẫn đến phần đánh giá thiếu khách quan”.

4.3. Thảo luận

Một yếu tố cần được cân nhắc kĩ lưỡng trong quá trình phân tích kết quả là đặc điểm của nhóm giám khảo tham gia đánh giá. Việc lựa chọn sinh viên trong nghiên cứu này có thể tạo ra một sự chênh lệch. Với nền tảng học vấn cao hơn, trải nghiệm học tập phong phú và mức độ thành thạo trong việc vận dụng tư duy phản biện, sinh viên có xu hướng đưa ra những đánh giá nhất quán và có cơ sở hơn, từ đó góp phần gia tăng độ tin cậy của đánh giá, khiến SSR = 0,87 trong bối cảnh này có thể không lặp lại khi triển khai với học sinh THPT. Do đó, kết quả về độ tin cậy và tác động học tập cần được kiểm chứng lại trên đối tượng đích trong những nghiên cứu sau.

Phân tích Infit ở mức sản phẩm cho thấy không có sản phẩm nào có Infit $\geq 1,3$; do đó, không quan sát thấy mức bất đồng đáng kể theo ngưỡng gợi ý. Tuy nhiên, một số sản phẩm có Infit cao tương đối (ví dụ: Nhóm 4: Infit $\approx 1,13$; Nhóm 8: Infit $\approx 1,11$) có thể phản ánh các trường hợp “biên” về chất lượng hoặc sự khác biệt trong cách giám khảo ưu tiên các khía cạnh. Cụ thể, ví dụ

ở nhóm 4, sản phẩm đạt mức độ chính xác cao về mặt tính toán tài chính (không có sai sót số học, các bảng tính được trình bày đầy đủ), nhưng chiến lược phân bổ danh mục đầu tư còn mang tính đơn giản, ít đa dạng hóa và chưa thể hiện rõ lập luận về cân bằng rủi ro và lợi nhuận. Trong quá trình đánh giá, một nhóm giám khảo có xu hướng ưu tiên tính chính xác và đầy đủ về mặt toán học nên thường đánh giá sản phẩm này cao hơn, trong khi nhóm giám khảo khác lại đặt trọng tâm vào tính hợp lý và sáng tạo của chiến lược đầu tư, dẫn đến các phán đoán trái chiều. Trường hợp này cho thấy giá trị infit cao có thể không phản ánh sự thiếu nhất quán ngẫu nhiên của giám khảo, mà xuất phát từ việc các giám khảo áp dụng những trọng số ưu tiên khác nhau đối với các khía cạnh chất lượng trong một nhiệm vụ học tập phức hợp. Kết quả này gợi ý rằng ở Bước 4 của quy trình, hoạt động đào tạo giám khảo cần làm rõ cách tiếp cận phán đoán toàn diện nhằm giúp sinh viên hình thành cách cân nhắc đồng thời nhiều khía cạnh khi thực hiện đánh giá so sánh.

Ngoài việc sử dụng công nghệ để tạo điều kiện thuận lợi cho quá trình đánh giá so sánh, một thuật toán ghép nối thích ứng các câu trả lời có thứ hạng tương tự được đề xuất để giảm thêm thời gian cần thiết để hoàn thành quá trình đánh giá, tuy nhiên, có nguy cơ làm sai lệch độ tin cậy của SSR theo hướng gia tăng không cần thiết. Các nghiên cứu tiếp theo có thể xem xét về việc sử dụng khả năng thích ứng và áp dụng ở mức độ nào khi đánh giá so sánh với số lượng lớn sản phẩm của học sinh trong đánh giá xác thực.

5. Kết luận

Nghiên cứu đã xây dựng quy trình 6 bước áp dụng đánh giá so sánh trong việc đánh giá các nhiệm vụ xác thực và tiến hành thử nghiệm sư phạm minh họa với sinh viên sư phạm toán trên một nhiệm vụ được thiết kế cho môn Toán lớp 11. Kết quả đánh giá so sánh cho thấy độ tin cậy tách thang đo đạt mức cao ($SSR = 0,87$) và các giá trị Infit đều dưới ngưỡng bất đồng đáng kể (1,3), đồng thời phản hồi định tính gợi ý người học dần hình thành cách diễn đạt rõ hơn về chất lượng của sản phẩm khi đưa ra phán đoán. Như vậy, phương pháp đánh giá so sánh có khả năng cải thiện độ tin cậy và tính khách quan trong việc chấm điểm, cũng là một phương thức tiềm năng trong việc tổ chức đánh giá như một hoạt động học tập. Điều này không chỉ phù hợp với định hướng đổi mới của Chương trình Giáo dục phổ thông 2018 mà còn góp phần nâng cao chất lượng dạy và học môn Toán hiện nay. Tuy nhiên, do đối tượng thử nghiệm là sinh viên, nghiên cứu thừa nhận hạn chế về khả năng ngoại suy và khuyến nghị cần tiếp tục triển khai, kiểm chứng và so sánh kết quả trên học sinh THPT trước khi đề xuất áp dụng rộng rãi.

Ghi chú về tác giả: Lê Thúy Diệu là giảng viên Khoa Toán – Tin, Trường Toán và Công nghệ thông tin, Trường Đại học Sư phạm Hà Nội. Phạm Hà Mai Trang là sinh viên tại Khoa Toán – Tin, Trường Đại học Sư phạm Hà Nội, Việt Nam. Tác giả 1: xây dựng ý tưởng, phương pháp nghiên cứu, giám sát nghiên cứu, viết bản thảo, chỉnh sửa và phản biện bản thảo; Tác giả 2: thu thập và phân tích dữ liệu, thử nghiệm sư phạm, trực quan hóa, viết bản thảo.

Tuyên bố về xung đột lợi ích. Các tác giả tuyên bố không có xung đột lợi ích.

TÀI LIỆU THAM KHẢO

- Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. *Educational Psychologist*, 48(2), 87–113. <https://doi.org/10.1080/00461520.2013.775712>
- Bartholomew, S., & Yoshikawa, E. (2018). A Systematic Review of Research Around Adaptive Comparative Judgement (ACJ) in K-16 Education. *CTETE - Research Monograph Series*, 1. <https://doi.org/10.21061/ctete-rms.v1.c.1>
- Bisson, M.-J., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring Conceptual Understanding Using Comparative Judgement. *International Journal of Research in Undergraduate Mathematics Education*, 2(2), 141–164. <https://doi.org/10.1007/s40753-016-0024-3>
- Bloxham, S., & Price, M. (2015). External examining: Fit for purpose? *Studies in Higher Education*, 40(2), 195–211. <https://doi.org/10.1080/03075079.2013.823931>

- Bouwer, R., Lesterhuis, M., Bonne, P., & De Maeyer, S. (2018). Applying Criteria to Examples or Learning by Comparison: Effects on Students' Evaluative Judgment and Performance in Writing. *Frontiers in Education*, 3. <https://doi.org/10.3389/educ.2018.00086>
- Cantley, I. (2023). Philosophical perspectives on authentic assessment in mathematics: The 15th International Congress on Mathematical Education. *15th International Congress on Mathematics Education. Topic Study Group: Proceedings, International Congress on Mathematics Education: Proceedings*.
- Ellison, C. (2023). Effects of adaptive comparative judgement on student engagement with peer formative feedback. *Practitioner Research in Higher Education*, 15(1), 24–35.
- Flores, M. A., Veiga Simão, A. M., Barros, A., & Pereira, D. (2015). Perceptions of effectiveness, fairness and feedback of assessment methods: A study in higher education. *Studies in Higher Education*, 40(9), 1523–1534. <https://doi.org/10.1080/03075079.2014.881348>
- Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education*, 39(10), 1774–1787. <https://doi.org/10.1080/03075079.2013.821974>
- Jones, I., & Inglis, M. (2015). The problem of assessing problem solving: Can comparative judgement help? *Educational Studies in Mathematics*, 89(3), 337–355. <https://doi.org/10.1007/s10649-015-9607-1>
- Jones, I., & Sirl, D. (2017). Peer assessment of mathematical understanding using comparative judgement. *NOMAD Nordic Studies in Mathematics Education*, 22(4), 147–164. <https://doi.org/10.7146/nomad.v22i4.148924>
- Jones, I., & Wheadon, C. (2015). Peer assessment using comparative and absolute judgment. *Studies in Educational Evaluation*, 47, 93–101. <https://doi.org/10.1016/j.stueduc.2015.09.004>
- Laming, D. (2003). *Human judgment: The eye of the beholder*. Cengage Learning EMEA.
- McConlogue, T. (2012). But is it fair? Developing students' understanding of grading complex written work through peer assessment. *Assessment & Evaluation in Higher Education*, 37(1), 113–123. <https://doi.org/10.1080/02602938.2010.515010>
- Mueller, J. (2005). *The Authentic Assessment Toolbox: Enhancing Student Learning through Online Faculty Development*. <https://www.semanticscholar.org/paper/The-Authentic-Assessment-Toolbox%3A-Enhancing-Student-Mueller/b4b5dfd76cb153ce12f303b66f8b4be1854c7f7b>
- Nguyễn, C. K., & Đào, T. O. (2015). Giáo trình kiểm tra đánh giá trong giáo dục. Hà Nội: NXB Đại học Sư phạm.
- Pachur, T., & Olsson, H. (2012). Type of learning task impacts performance and strategy selection in decision making. *Cognitive Psychology*, 65(2), 207–240. <https://doi.org/10.1016/j.cogpsych.2012.03.003>
- Pollitt, A. (2012). The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice*, 19(3), 281–300. <https://doi.org/10.1080/0969594X.2012.665354>
- Pollitt, A., & Whitehouse, C. (2012). *Using adaptive comparative judgement to obtain a highly reliable rank order in summative assessment*.
- 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>
- Sadler, D. R. (2009). Indeterminacy in the Use of Preset Criteria for Assessment and Grading. *Assessment & Evaluation in Higher Education*, 34(2), 159–179. <https://doi.org/10.1080/02602930801956059>
- Seery, N., Canty, D., & Phelan, P. (2012). The validity and value of peer assessment using adaptive comparative judgement in design driven practical education. *International Journal of Technology and Design Education*, 22(2), 205–226. <https://doi.org/10.1007/s10798-011-9194-0>
- Svinicki, M. D. (2004). Authentic assessment: Testing in reality. *New Directions for Teaching and Learning*, 2004(100), 23–29. <https://doi.org/10.1002/tl.167>
- Đặng, T.T.T (2022). Đánh giá như hoạt động học tập và học tập tự định hướng ở các cơ sở giáo dục đại học. *Tạp chí Khoa học Giáo dục Việt Nam*, 02(2), 7–12.
- van Daal, T., Lesterhuis, M., Coertjens, L., Donche, V., & De Maeyer, S. (2019). Validity of comparative judgement to assess academic writing: Examining implications of its holistic character and building on a shared consensus. *Assessment in Education: Principles, Policy & Practice*, 26(1), 59–74. <https://doi.org/10.1080/0969594X.2016.1253542>
- van Daal, T., Snajder, M., Nijs, K., & Van Dyck, H. (2023). Peer Assessment Using Criteria or Comparative Judgement? A Replication Study on the Learning Effect of Two Peer Assessment Methods. Trong O. Noroozi & B. De Wever (B.t.v.), *The Power of Peer Learning: Fostering Students' Learning Processes and Outcomes* (tr 73–101). Springer International Publishing. https://doi.org/10.1007/978-3-031-29411-2_4
- Wilson, J. W. (1973). Standardized Tests Very Often Measure the Wrong Things. *The Mathematics Teacher*, 66(4), 295–370.