

Bước đầu xây dựng rubric phân tích đánh giá giao tiếp nói tiếng Anh cho sinh viên không chuyên

Ngày nhận bài: 07/08/2026 | Ngày phản biện: 18/08/2026 | Ngày xuất bản: 22/09/2026

Tác giả: Lê Thị Thanh Trang (Học viện Hành chính và Quản trị công)

Tóm tắt: Bài viết trình bày giai đoạn đầu xây dựng rubric phân tích đánh giá năng lực giao tiếp nói tiếng Anh của sinh viên không chuyên trong học phần Tiếng Anh giao tiếp và hội nhập. Nghiên cứu theo hướng lý luận – ứng dụng, dựa trên ba sách chuyên khảo và hai tài liệu gần đây về kiểm tra ngôn ngữ, thiết kế đánh giá. Từ các nguyên tắc về cấu trúc năng lực, độ giá trị, tính khả thi, thiết kế nhiệm vụ và thang điểm, nghiên cứu xác định năm tiêu chí: hoàn thành nhiệm vụ; tổ chức diễn ngôn và phù hợp ngữ dụng; nguồn lực ngôn ngữ; độ trôi chảy và độ dễ hiểu; tương tác và chiến lược giao tiếp. Rubric gồm bốn mức từ 1 đến 4, hướng tới chuẩn đầu ra A2+ đến B1-; điểm 0 dùng khi không đủ minh chứng. Bài viết đề xuất hai phần nhiệm vụ khoảng 10 phút cho một cặp, cách tính điểm và lộ trình thẩm định bằng chuyên gia, thử nghiệm nhiệm vụ, kiểm tra mức độ nhất quán giữa người chấm. Công cụ cần được kiểm chứng trước khi sử dụng.

1. Đặt vấn đề

Ở bậc đại học, các học phần tiếng Anh cho sinh viên không chuyên ngày càng chú trọng khả năng sử dụng ngôn ngữ. Sinh viên không chỉ cần nhớ từ vựng và ngữ pháp, mà còn phải dùng tiếng Anh để trình bày ý tưởng, trao đổi thông tin, hợp tác và xử lý các tình huống học tập hoặc nghề nghiệp cơ bản. Vì vậy, bài kiểm tra cần cho thấy sinh viên thực sự giao tiếp như thế nào, thay vì chỉ đo khả năng nhận biết kiến thức ngôn ngữ.

Đánh giá nói thường được tổ chức qua thuyết trình, phỏng vấn, trao đổi theo cặp, đóng vai hoặc giải quyết vấn đề. Theo Hughes và Hughes, nhiệm vụ nói cần giống với hoạt động giao tiếp mà người học sẽ phải thực hiện và phải tạo ra đủ biểu hiện để người chấm đưa ra nhận định hợp lý (Hughes & Hughes, 2020, tr. 113). Nếu nhiệm vụ không gắn với chuẩn đầu ra, bài nói có thể không phản ánh đúng năng lực cần đánh giá. Nếu tiêu chí không rõ, giảng viên dễ chấm theo ấn tượng chung hoặc chú ý quá nhiều đến một vài mặt của bài nói.

Rubric phân tích là bảng chấm tách bài nói thành nhiều tiêu chí và cho điểm từng tiêu chí. Cách chấm này vừa cho điểm, vừa cho biết người học làm tốt ở đâu và cần cải thiện điều gì. Brown và Abeywickrama cho rằng rubric phân tích giúp phản hồi cụ thể hơn cho từng mặt của bài nói (Brown & Abeywickrama, 2019, tr. 251). Một công cụ đánh giá trên lớp cũng cần cân bằng năm yêu cầu: dễ tổ chức, kết quả ổn định, đo đúng năng lực cần đo, gần với giao tiếp thực tế và có tác động tích cực đến việc học (Brown & Abeywickrama, 2019, tr. 27). Vì vậy, mục đích đánh giá, nhiệm vụ, tiêu chí, minh chứng và cách hiểu điểm số phải gắn kết với nhau.

Tiếng Anh giao tiếp và hội nhập là học phần bắt buộc gồm 5 tín chỉ, chủ yếu dành cho sinh viên năm thứ nhất không chuyên tiếng Anh. Học phần có 45 tiết lý thuyết và 60 tiết thực hành; ngôn ngữ giảng dạy là tiếng Anh. Chuẩn đầu ra ở khoảng A2+ đến B1-. Sinh viên phát triển bốn kỹ năng qua các chủ đề quen thuộc như ẩm thực, khát vọng, du lịch, nghề nghiệp, giải trí, công nghệ và thách thức toàn cầu. Ở kỹ năng nói, sinh viên cần biết trình bày quan điểm, thuyết phục, từ chối, diễn đạt lại, đặt câu hỏi mở và phản hồi tích cực. Do đó, rubric phải đánh giá được nội dung, cách tổ chức ý, ngôn ngữ, độ trôi chảy, khả năng tương tác và cách xử lý khó khăn. Đồng thời, công cụ phải đủ gọn để dùng trong lớp đông.

Bài viết tập trung vào ba việc: xác định các tiêu chí cốt lõi, mô tả bốn mức thực hiện và đề xuất cách sử dụng cùng lộ trình thẩm định rubric.

2. Phương pháp nghiên cứu

Nghiên cứu sử dụng cách tiếp cận định tính, gồm phân tích tài liệu, nhóm nội dung theo chủ đề, đối chiếu các khái niệm và tiêu chí, sau đó xây dựng công cụ ban đầu. Đối tượng phân tích là cơ sở lý luận về rubric đánh giá giao tiếp nói và các yêu cầu

đánh giá trong học phần Tiếng Anh giao tiếp và hội nhập. Nguồn phân tích gồm ba sách chuyên khảo, hai nghiên cứu gần đây về rubric đánh giá nói và đề cương học phần. Một tiêu chí chỉ được chọn khi có cơ sở lý luận, có thể quan sát trong nhiệm vụ nói và có thể sử dụng trong lớp học. Nghiên cứu chưa khảo sát người học, chưa thu thập bài nói và chưa kiểm định bằng dữ liệu thực tế. Vì vậy, kết quả mới là đề xuất ban đầu.

3. Kết quả nghiên cứu và thảo luận

3.1. Cơ sở lý luận về xây dựng rubric đánh giá năng lực giao tiếp nói tiếng

3.1.1. Rubric phân tích và các yêu cầu đối với công cụ đánh giá

Trong rubric phân tích, giảng viên không chỉ cho một điểm chung cho toàn bài nói. Bài thực hiện được chia thành các tiêu chí và mỗi tiêu chí có một điểm riêng. Brown và Abeywickrama cho rằng cách này tạo ra phản hồi cụ thể hơn, nhưng người thiết kế phải xác định rõ từng thành phần và từng mức để tránh các tiêu chí bị trùng nhau (Brown & Abeywickrama, 2019, tr. 251). Vì thế, rubric đánh giá nói cần mô tả những biểu hiện có thể nghe hoặc quan sát được, thay vì chỉ liệt kê kiến thức về từ vựng và ngữ pháp.

Fulcher và Davidson cho rằng mô tả mức điểm phải xuất phát từ những thành phần của năng lực giao tiếp phù hợp với mục đích và phạm vi sử dụng của bài đánh giá (Fulcher & Davidson, 2007, tr. 93). Nói cách khác, chuẩn đầu ra, nhiệm vụ, tiêu chí, mô tả mức và cách hiểu điểm số phải thống nhất. Mô tả quá chung khiến người chấm khó chọn mức phù hợp với bài nói thực tế. Ngược lại, mô tả quá dài và chi tiết sẽ khó sử dụng khi chấm nhiều sinh viên.

Brown và Abeywickrama (2019) đề xuất năm nguyên tắc để xem xét chất lượng của công cụ đánh giá trên lớp. Tính khả thi cho biết công cụ có phù hợp với thời gian và điều kiện tổ chức hay không. Độ tin cậy thể hiện ở việc các lần chấm hoặc những người chấm cho kết quả tương đối ổn định. Độ giá trị cho biết nhiệm vụ và tiêu chí có đo đúng năng lực cần đo hay không. Tính xác thực là mức độ nhiệm vụ gần với giao tiếp thực tế. Tác động đối với việc học thể hiện ở việc rubric có giúp người học luyện tập và nhận phản hồi tốt hơn hay không.

3.1.2. Cấu trúc năng lực giao tiếp nói và căn cứ lựa chọn tiêu chí

Trong bài viết này, năng lực giao tiếp nói là khả năng kết hợp ngôn ngữ với cách tổ chức ý, cách dùng lời phù hợp bối cảnh, khả năng tương tác và chiến lược xử lý khó khăn để đạt mục đích giao tiếp. Fulcher và Davidson lưu ý rằng năng lực không thể được nhìn thấy trực tiếp. Vì vậy, người thiết kế phải chuyển năng lực đó thành những biểu hiện có thể quan sát và đo được (Fulcher & Davidson, 2007, tr.293). Mỗi tiêu chí trong rubric do đó phải gắn với dấu hiệu cụ thể trong lời nói và sự tương tác của sinh viên.

Ogawa xem xét bài nói qua cách tổ chức nội dung, mức độ phức tạp, độ chính xác và độ trôi chảy. Tác giả cũng nhấn mạnh rằng bài đánh giá giao tiếp phải xem người nói có đạt được mục đích giao tiếp hay không (Ogawa, 2022, tr.1). Điều này cho thấy độ chính xác của ngôn ngữ chỉ là một phần của bài nói. Nội dung, cách sắp xếp thông điệp và hiệu quả giao tiếp cũng cần được đánh giá.

Đorđević nghiên cứu nhiều loại rubric, gồm rubric mô tả, rubric phân tích, rubric theo nhiệm vụ và rubric do giảng viên xây dựng. Các loại này đều có thang điểm, tiêu chí, các mức thành tích và mô tả cho từng mức (Đorđević, 2025, tr.91). Kết quả cũng cho thấy rubric có thể giúp người học chuẩn bị nhiệm vụ và giúp giảng viên đánh giá bài nói, nếu tiêu chí được viết rõ và gắn với cách dùng ngôn ngữ trong tình huống thực (Đorđević, 2025, tr.106).

3.2. Căn cứ và bước đầu xây dựng rubric

3.2.1. Yêu cầu đánh giá trong học phần

Học phần Tiếng Anh giao tiếp và hội nhập đặt chuẩn đầu ra ở mức A2+ đến B1- cho sinh viên không chuyên. Trong phần nói, sinh viên cần trình bày quan điểm, trao đổi thông tin, thuyết phục, từ chối, đặt câu hỏi mở, phản hồi và diễn đạt lại. Vì vậy, bài đánh giá phải có cả phần nói cá nhân ngắn và phần tương tác với người khác. Chỉ kiểm tra riêng phát âm, từ vựng hoặc ngữ pháp sẽ không tạo đủ minh chứng về năng lực giao tiếp.

Phân tích đề cương học phần cho thấy chuẩn đầu ra cần được chuyển thành những tiêu chí và hành vi có thể quan sát. Nếu chỉ cho một điểm chung hoặc dùng các nhãn mức quá rộng, giảng viên khó giải thích vì sao sinh viên nhận điểm đó. Sinh viên cũng không biết mình cần cải thiện ở mặt nào. Tuy nhiên, do lớp đông và thời gian vấn đáp

có hạn, rubric phải đủ ngắn và dễ dùng.

3.2.2. Các tiêu chí và mức thực hiện được đề xuất

Từ kết quả phân tích tài liệu, nghiên cứu đề xuất năm tiêu chí. Tiêu chí 1, hoàn thành nhiệm vụ giao tiếp, xem sinh viên có đạt đúng mục đích, chọn nội dung phù hợp và điều chỉnh thông tin cho người nghe hay không. Tiêu chí 2, tổ chức diễn ngôn và phù hợp ngữ dụng, xem cách phát triển và liên kết ý, đồng thời xem cách đề nghị, giải thích, phản hồi hoặc thuyết phục có phù hợp với vai trò và bối cảnh hay không. Tiêu chí 3, nguồn lực ngôn ngữ, xem phạm vi và mức kiểm soát từ vựng, khả năng kết hợp câu đơn, câu ghép và mệnh đề phụ đơn giản, cùng mức độ lỗi cản trở giao tiếp. Tiêu chí 4, độ trôi chảy và độ dễ hiểu, xem khả năng duy trì dòng lời, ngắt nghỉ, phát âm, trọng âm, nhịp điệu và mức độ người nghe hiểu thông điệp. Tiêu chí 5, tương tác và chiến lược giao tiếp, xem khả năng phản hồi, quản lý lượt lời, hỏi lại, làm rõ, diễn đạt lại và tự sửa.

Rubric có bốn mức từ 1 đến 4 và sử dụng cùng một nhóm dấu hiệu ở tất cả các mức. Mức 4 cho thấy sinh viên thực hiện tốt và khá độc lập trong phạm vi chủ đề và chức năng giao tiếp của chuẩn A2+ đến B1-. Mức 3 cho thấy sinh viên đạt mục đích chính nhưng vẫn còn một số hạn chế nhỏ. Mức 2 cho thấy sinh viên chỉ hoàn thành một phần và thường cần hỗ trợ. Mức 1 cho thấy sinh viên chưa đạt mục đích hoặc tạo ra rất ít minh chứng. Điểm 0 không phải là một mức năng lực. Điểm này chỉ dùng khi sinh viên không thực hiện nhiệm vụ hoặc không tạo đủ minh chứng để chấm. Tổng điểm cao nhất là 20. Điểm thang 10 bằng tổng điểm chia 2; điểm thang 100 bằng tổng điểm nhân 5.

3.2.3. Nhiệm vụ và cách tổng hợp minh chứng

Để phù hợp với thời lượng khoảng 10 phút cho một cặp, bài đánh giá gồm hai phần. Phần trả lời cá nhân kéo dài khoảng 4 phút cho cả cặp, tức mỗi sinh viên khoảng 2 phút. Phần này chủ yếu cho thấy mức độ hoàn thành nhiệm vụ, cách tổ chức thông điệp, nguồn lực ngôn ngữ, độ trôi chảy và độ dễ hiểu. Phần giải quyết vấn đề theo cặp kéo dài khoảng 5-6 phút. Đây là nguồn minh chứng chính cho khả năng tương tác và sử dụng chiến lược, đồng thời tiếp tục cung cấp minh chứng cho bốn tiêu chí còn lại. Cách kết hợp này phù hợp với yêu cầu nhiệm vụ nói phải tạo ra hành vi đại diện cho

năng lực cần đánh giá (Hughes & Hughes, 2020, tr.113).

Hai phần nhiệm vụ không có điểm và tỷ trọng riêng. Người chấm xem minh chứng từ cả hai phần rồi cho một điểm từ 0 đến 4 ở mỗi tiêu chí. Năm tiêu chí có trọng số bằng nhau, mỗi tiêu chí chiếm 20% tổng điểm. Tiêu chí 5 chủ yếu được chấm từ phần tương tác theo cặp. Nếu không thể tổ chức theo cặp, nhiệm vụ thay thế vẫn phải có sự trao đổi thực. Không nên suy ra năng lực tương tác chỉ từ một bài nói cá nhân.

3.3. Ưu điểm của rubric đề xuất

Ưu điểm thứ nhất là rubric có phạm vi đánh giá rộng. Ngoài từ vựng, ngữ pháp và phát âm, công cụ còn xem xét mục đích giao tiếp, cách tổ chức ý, cách dùng lời phù hợp bối cảnh, khả năng tương tác và chiến lược xử lý khó khăn. Cách lựa chọn này phù hợp với quan điểm của Ogawa về việc đánh giá nội dung, độ phức tạp, độ chính xác, độ trôi chảy và mức độ đạt mục đích giao tiếp (Ogawa, 2022, tr.1).

Ưu điểm thứ hai là rubric nổi chuẩn đầu ra với nhiệm vụ và những biểu hiện có thể quan sát. Năm tiêu chí dùng cùng một hệ thống dấu hiệu qua bốn mức. Nhờ đó, người chấm có thể dựa vào những gì sinh viên thực sự làm trong bài nói, thay vì chủ yếu dựa vào ấn tượng chung. Điểm số vì thế được hiểu trong đúng phạm vi của nhiệm vụ đã thực hiện.

Ưu điểm thứ ba là công cụ tương đối phù hợp với lớp đông. Trong rubric, cách tổ chức ý được gộp với cách dùng lời phù hợp bối cảnh; tương tác được gộp với chiến lược giao tiếp. Nhờ vậy, rubric chỉ có năm tiêu chí. Số lượng này vẫn bao quát các mặt chính của bài nói nhưng không làm phiếu chấm quá dài.

Ưu điểm thứ tư là rubric có thể hỗ trợ cả việc chấm điểm và việc học. Khi nhận rubric trước, sinh viên có thể dùng tiêu chí để chuẩn bị, tự đánh giá hoặc nhận xét bài của bạn học. Đorđević cho thấy một rubric cụ thể có thể vừa hướng dẫn người học chuẩn bị, vừa giúp giảng viên đánh giá hoạt động nói (Đorđević, 2025, tr.106). Phản hồi theo từng tiêu chí cũng giúp giảng viên nêu rõ một điểm mạnh và một việc cần ưu tiên cải thiện, thay vì chỉ thông báo điểm tổng.

3.4. Hạn chế, khó khăn và nguyên nhân

3.4.1. Hạn chế

Hạn chế lớn nhất là rubric mới được xây dựng từ tài liệu lý luận và yêu cầu của học phần. Công cụ chưa được chuyên gia thẩm định, chưa được thử với bài nói thực tế và chưa có kết quả chấm độc lập của nhiều người chấm. Vì vậy, nghiên cứu chưa thể khẳng định rubric đã đo đúng năng lực cần đánh giá hoặc đã cho kết quả ổn định giữa các giảng viên.

Một hạn chế khác nằm ở việc gộp các thành phần. Khi tổ chức diễn ngôn được gộp với phù hợp ngữ dụng, điểm số có thể không cho biết rõ sinh viên yếu ở cách sắp xếp thông điệp hay ở cách dùng lời phù hợp bối cảnh. Tương tự, việc gộp tương tác với chiến lược giao tiếp có thể làm khó việc tìm nguyên nhân của điểm thấp. Ngoài ra, bốn mức điểm có thể tạo ra các trường hợp nằm giữa hai mức và khiến điểm tập trung nhiều ở mức 2 hoặc 3.

3.4.2. Khó khăn

Thứ nhất, phải tạo đủ minh chứng cho cả năm tiêu chí trong khoảng 10 phút. Nếu câu hỏi cá nhân quá dài, phần trao đổi theo cặp sẽ bị rút ngắn. Nếu nhiệm vụ cặp không buộc hai sinh viên phải trao đổi thật, người chấm sẽ không có đủ căn cứ cho tiêu chí tương tác và chiến lược.

Thứ hai, giữ cách chấm thống nhất. Một biểu hiện như ngắt nghỉ, tự sửa, lỗi ngôn ngữ hoặc thiếu nội dung có thể liên quan đến nhiều tiêu chí. Nếu không có hướng dẫn rõ, người chấm có thể trừ điểm cùng một lỗi nhiều lần. Người chấm cũng có thể bị hiệu ứng hào quang, nghĩa là ấn tượng chung về bài nói ảnh hưởng đến tất cả các tiêu chí, hoặc quá chú ý đến ngữ pháp và phát âm.

Thứ ba, tổ chức đánh giá trong lớp đông. Việc cho sinh viên thi theo cặp, ghi minh chứng và phản hồi từng tiêu chí cần nhiều thời gian, nhân lực và một lịch kiểm tra phù hợp. Việc ghi âm hoặc ghi hình để kiểm tra lại kết quả còn phụ thuộc vào quy định và điều kiện kỹ thuật của cơ sở đào tạo.

3.4.3. Nguyên nhân

Nguyên nhân chính là năng lực giao tiếp nói gồm nhiều thành phần. Kết quả của một bài nói ngắn có thể bị ảnh hưởng bởi chủ đề, dạng nhiệm vụ, người cùng trao đổi, tâm lý của sinh viên và cách giảng viên điều hành bài thi. Vì vậy, điểm của một nhiệm vụ ngắn không thể đại diện cho toàn bộ năng lực tiếng Anh của người học.

Nguyên nhân tiếp theo là các mô tả mức hiện chủ yếu được xây dựng từ lý thuyết, chưa được điều chỉnh bằng bài nói của đúng nhóm sinh viên mục tiêu. Fulcher và Davidson lưu ý rằng mô tả quá chung khiến người chấm khó ghép minh chứng cụ thể với mức điểm (Fulcher & Davidson, 2007, tr. 95). Bên cạnh đó, khi chưa có bài mẫu chuẩn và chưa tập huấn chung, mỗi người chấm có thể hiểu tiêu chí theo một cách khác nhau.

3.5. Giải pháp hoàn thiện và sử dụng rubric

Thứ nhất, mời 5-7 chuyên gia hoặc giảng viên có kinh nghiệm thẩm định nội dung. Họ cần xem tiêu chí có phù hợp với chuẩn đầu ra hay không, thuật ngữ có rõ không, các tiêu chí có bị chồng lấn không, bốn mức có phân biệt được với nhau không và công cụ có thể dùng trong lớp học hay không. Ý kiến thẩm định phải được dùng để sửa trực tiếp rubric, thay vì chỉ được tổng hợp thành nhận xét.

Thứ hai, thử nghiệm nhiệm vụ với khoảng 30-50 bài nói của đúng nhóm sinh viên mục tiêu. Cần thu thập điểm từng tiêu chí, thời gian thực hiện và thời gian chấm, những trường hợp không có đủ minh chứng và phản hồi của người học. Cũng cần xem có quá nhiều sinh viên nhận điểm thấp nhất hoặc cao nhất hay không. Nếu một tiêu chí thường xuyên thiếu minh chứng hoặc dễ bị nhầm với tiêu chí khác, nhiệm vụ và rubric phải được sửa.

Thứ ba, viết hướng dẫn chấm và tập huấn người chấm bằng các bài nói mẫu. Người chấm nên nghe hoặc xem toàn bộ bài thực hiện trước khi cho điểm từng tiêu chí. Họ cần ghi minh chứng ngắn gọn, tránh trừ một lỗi ở nhiều tiêu chí và cùng thảo luận những trường hợp nằm giữa hai mức. Với bài đánh giá có hệ số cao, hai người chấm nên làm việc độc lập trên ít nhất 20% số bài nếu điều kiện cho phép. Trường hợp hai người chênh nhau hai mức ở một tiêu chí phải được xem lại.

Thứ tư, kiểm tra sự thống nhất giữa người chấm và xem từng tiêu chí có hoạt động đúng như dự kiến hay không. Nghiên cứu của Đorđević cho thấy có thể dùng dữ liệu từ nhiều đợt chấm và nhiều người chấm để xem xét độ giá trị, độ tin cậy và khả năng lặp lại của rubric (Đorđević, 2025, tr. 102-105). Kết quả sẽ giúp quyết định giữ trọng số bằng nhau, thay đổi trọng số hoặc tách các tiêu chí đang được gộp. Sau khi sửa, rubric cần được thử lại trên một nhóm nhỏ trước khi ban hành hướng dẫn sử dụng.

Thứ năm, cung cấp rubric cho sinh viên trước khi đánh giá. Giảng viên cần giải thích mỗi tiêu chí bằng ví dụ ngắn và cho sinh viên luyện tự đánh giá. Sau bài nói, phản hồi nên gồm điểm của từng tiêu chí, một điểm mạnh và một việc cần ưu tiên cải thiện. Với lớp đông, giảng viên có thể dùng mã phản hồi ngắn hoặc chỉ chọn một tiêu chí trọng tâm trong mỗi lần đánh giá quá trình.

4. Kết luận

Bài viết đã đề xuất bước đầu một rubric phân tích để đánh giá năng lực giao tiếp nói tiếng Anh của sinh viên không chuyên trong học phần Tiếng Anh giao tiếp và hội nhập. Rubric có năm tiêu chí và bốn mức từ 1 đến 4, hướng tới chuẩn A2+ đến B1-. Điểm 0 chỉ dùng khi không có đủ minh chứng để chấm.

Rubric được sử dụng với hai phần nhiệm vụ: trả lời cá nhân và giải quyết vấn đề theo cặp. Tổng thời gian khoảng 10 phút cho một cặp. Người chấm kết hợp minh chứng từ cả hai phần để cho một điểm ở mỗi tiêu chí. Năm tiêu chí có trọng số bằng nhau.

Điểm mạnh của công cụ là đánh giá được mục đích giao tiếp, cách tổ chức ý, cách dùng lời phù hợp bối cảnh, nguồn lực ngôn ngữ, độ trôi chảy, khả năng tương tác và chiến lược giao tiếp trong một rubric tương đối gọn. Tuy nhiên, đây mới là phiên bản ban đầu. Trước khi dùng cho đánh giá cuối học phần, rubric cần được chuyên gia thẩm định, thử với bài nói thực tế và kiểm tra sự thống nhất giữa người chấm. Dữ liệu thử nghiệm sẽ giúp điều chỉnh nhiệm vụ, mô tả mức, trọng số hoặc quyết định có cần tách tiêu chí ở phiên bản sau hay không.

Tài liệu tham khảo

- Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.
- Dorđević, J. (2025). Rubrics in the assessment of EAP speaking skills supported by mobile assisted language learning. *ESP Today*, 13(1), 91-112. <https://doi.org/10.18485/esptoday.2025.13.1.5>
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge.
- Hughes, A., & Hughes, J. (2020). *Testing for language teachers* (3rd ed.). Cambridge University Press.
- Ogawa, C. (2022). CAF indices and human ratings of oral performances in an opinion-based monologue task. *Language Testing in Asia*, 12, Article 4. <https://doi.org/10.1186/s40468-022-00154-9>