

## NHẬN THỨC CỦA GIÁNG VIÊN VỀ HAI THANG ĐO ĐÁNH GIÁ KỸ NĂNG NÓI TIẾNG ANH: THANG ĐO TỔNG QUÁT VÀ THANG ĐO CỤ THỂ

Lương Trọng Linh và các cộng sự<sup>54</sup>

**Tóm tắt:** Trong việc đánh giá kỹ năng nói của người học, giảng viên thường hay sử dụng hai thang đo đánh giá chủ yếu: Thang đo tổng quát (holistic rubric) và thang đo cụ thể (analytic rubric). Để hiểu rõ hơn về nhận thức của giảng viên về hai thang đo này, 42 câu hỏi được thiết kế theo thang đo Likert được sử dụng để phân tích với sự tham gia trả lời của 40 giảng viên tiếng Anh. Thêm vào đó, 9 giảng viên được chọn lựa ngẫu nhiên sẽ trả lời câu hỏi phỏng vấn để có cái nhìn chi tiết về những loại thang đo mà họ thích và xem nhận thức của họ được phản ánh trong việc thực hành đánh giá người học dựa trên 2 thang đo này như thế nào. Kết quả cho thấy hầu hết các giảng viên có nhận thức tốt về tầm quan trọng của việc sử dụng các thang đo đánh giá trong việc đánh giá kỹ năng nói, về việc sử dụng thang đo tổng quát (holistic rubric) và thang đo cụ thể (analytic rubric) trong đánh giá kỹ năng nói của người học. Bên cạnh đó, trong việc đánh giá kỹ năng nói, giảng viên thường chọn mỗi loại thang đo đánh giá riêng dựa vào những lợi ích mà mỗi thang đo đánh giá mang lại hoặc giáo viên sẽ chọn kết hợp cả 2 thang đo cùng một lúc để đánh giá kỹ năng nói của người học nhằm mang lại độ tin cậy cao hơn.

**Từ khóa:** Nhận thức của giảng viên, kỹ năng nói tiếng Anh, thang đo tổng quát, thang đo cụ thể, đánh giá kỹ năng nói.

**Abstract:** In assessing speaking skill, lecturers usually employ two typical rubrics: holistic rubric and analytic rubric. To find out lecturer's perceptions of holistic and analytic rubrics in assessing speaking, a questionnaire was designed and used to collect data from 40 university lecturers. In addition, nine of the participants were then asked to sit for the in-depth interview in order to obtain more insightful understanding of the type of rubric they preferred and of why and how they used holistic and analytic rubrics in assessing speaking. The results have shown that most of the investigated lecturers had good perceptions of the significance of using rubrics in assessing speaking, of using holistic and analytic rubrics in assessing the student's speaking performance. Besides that, in assessing speaking, the lecturers tended to choose a specific type of rubric based on its benefits, or to combine both holistic and analytic rubrics in assessing speaking skill in order to bring about a high level of reliability.

**Key words:** lecturer's perceptions, holistic rubric, analytic rubric, assessing speaking

### 1. ĐẶT VẤN ĐỀ

Theo phát biểu của Ur (1996), kỹ năng nói đòi hỏi người sử dụng ngôn ngữ phải kết hợp nhiều khía cạnh khác của ngôn ngữ như ngữ âm, ngữ điệu, từ vựng, ngữ pháp, ngữ cảnh và đồng thời diễn tả ý nghĩ của người nói vì thế mà ông cho rằng kỹ năng nói tiếng Anh bao hàm

<sup>54</sup> Trần Thị Thùy, Dương Minh Tuấn, Phạm Đình Quốc, Nguyễn Hai Khoa, Trần Thanh Nga, Lê Trung Kiên, Đặng Vũ Hoài Nhân, Đào Thị Thái, Nguyễn Huỳnh Huế Thư, Lương Tú Ngân, Lê Thị Diễm Lan - Trường Đại học Nam Cần Thơ

các kỹ năng còn lại. Cũng chính vì vậy mà kỹ năng nói tiếng Anh được xem là kỹ năng quan trọng nhất khi học một ngôn ngữ nước ngoài. Tuy nhiên, bằng cách nào để đánh giá kỹ năng nói tiếng Anh từ lâu đã trở thành câu hỏi khó vì kỹ năng nói tiếng Anh được xem như là kỹ năng khó đánh giá nhất so với các kỹ năng nghe, đọc hay viết. Lý do hàng đầu theo Luoma (2004) là vì kỹ năng nói là một kỹ năng giao tiếp đòi hỏi sự tương tác thay đổi liên tục và diễn ra đồng thời.

Trong việc đánh giá kỹ năng nói, có những tranh luận liên quan đến các loại đánh giá. Đó chính là hình thức đánh giá dựa trên cái tổng quát và hình thức đánh giá dựa trên các tiêu chí cụ thể. Theo Huot (1990), hình thức đánh giá tổng quát mang tính linh hoạt, tiết kiệm và thực tế hơn, nhưng theo Brown (2004) lại cho rằng hình thức đánh giá dựa trên các tiêu chí cụ thể mang lại nhiều lợi ích hơn, cụ thể là hình thức đánh giá này có thể đánh giá chi tiết hơn các khía cạnh của kỹ năng nói như là tính chính xác, độ thành thạo, ngữ âm, ngữ điệu...

Các tranh luận này cho thấy rằng các giảng viên có những nhận thức khác nhau về hai cách đánh giá kỹ năng nói tiếng Anh, cách đánh giá tổng quát và cách đánh giá dựa trên chi tiết, cụ thể. Việc tìm ra nhận thức và hành động của họ giúp hiểu rõ nhận thức của giảng viên thực tại và đề ra các giải pháp hữu ích cho việc đánh giá kỹ năng nói tiếng Anh. Cũng chính vì những lý do đó mà đề tài **“Nhận thức của giảng viên về hai thang đo đánh giá trong tiếng Anh: thang đo đánh giá tổng quát và thang đo đánh giá cụ thể”** được thực hiện.

## 2. MỤC ĐÍCH NGHIÊN CỨU

Nghiên cứu này nhằm để làm rõ những nhận thức của giảng viên về thang đo tổng quát và thang đo cụ thể trong việc đánh giá kỹ năng nói tiếng Anh. Bên cạnh đó, nghiên cứu này hướng đến việc tìm hiểu nhận thức của họ về hai thang đo này được phản ánh qua việc thực hiện của họ như thế nào khi họ đánh giá kỹ năng nói tiếng Anh. Từ đó đề ra những giải pháp sư phạm phù hợp nhằm nâng cao chất lượng học tập.

## 3. THANG ĐO TỔNG QUÁT VÀ THANG ĐO CỤ THỂ

Theo MCNamara (2000), thang đo đánh giá cụ thể dựa trên việc “đánh giá từng khía cạnh một cách riêng biệt”. Do đó, đánh giá khả năng nói theo thang đo này yêu cầu đánh giá theo từng thành phần như phát âm, từ vựng, độ trôi chảy, ngữ pháp, nội dung...v.v. (Alderson và cộng sự, 1995). Alderson, Clapham và Wall (1995) giải thích rằng đánh giá theo thang đo cụ thể là phân tích sự thể hiện của thí sinh về các thành phần khác nhau. Kết quả của việc đánh giá dựa trên nhiều thành phần này, một điểm tổng thể xuất hiện ở cuối quy trình đánh giá. Điểm tổng thể theo thang đo cụ thể được tính bằng cách cộng tất cả các điểm của các thành phần đánh giá khác nhau (Alderson, Clapham & Wall, 1995).

Thang đo đánh giá tổng quát, còn được gọi là thang đo đánh giá truyền thống (Lumley, 2005), tập trung vào tổng thể (Brown, 2004). Trong kiểu đánh giá này, kỹ năng được kiểm tra được coi là một thể thống nhất, không có khả năng đánh giá từng phần một cách độc lập với phần kia. Alderson, Clapham và Wall (1995) nói rằng trong thang đo đánh giá tổng quát, “người chấm thi được yêu cầu không chú ý quá nhiều đến bất kỳ khía cạnh nào của ứng viên mà là đánh giá hiệu quả tổng thể của nó”. Trong thang đo đánh giá tổng quát về khả năng nói, người kiểm tra sẽ xem xét sự thể hiện kỹ năng nói chung. Các thành phần khác nhau, cụ thể là độ trôi chảy, từ vựng, phát âm, ngữ pháp và khả năng hiểu, được đánh giá tổng thể.

## 4. PHƯƠNG PHÁP NGHIÊN CỨU

### 4.1 Câu hỏi nghiên cứu

Để tìm ra nhận thức của giảng viên về hai thang đo đánh giá kỹ năng nói tiếng Anh, hai câu hỏi nghiên cứu được đặt ra.

**Câu hỏi nghiên cứu 1:** Nhận thức của giảng viên về các thang đo đánh giá tổng quát và thang đo đánh giá cụ thể trong việc đánh giá khả năng nói?

**Câu hỏi nghiên cứu 2:** Nhận thức của giảng viên được phản ánh như thế nào trong thực hành sử dụng thang đo đánh giá tổng quát và cụ thể trong đánh giá kỹ năng nói?

### 4.2 Ứng viên

40 giảng viên trả lời các câu hỏi từ bảng câu hỏi và 9 giảng viên được chọn ngẫu nhiên phỏng vấn trực tiếp. Tất cả đều là giảng viên tiếng Anh có bằng đại học hoặc thạc sỹ tại địa bàn thành phố Cần Thơ. Tất cả họ đều có ít nhất 2 năm kinh nghiệm giảng dạy và đã tham gia lớp đào tạo nghiệp vụ sư phạm.

### 4.3 Công cụ nghiên cứu

**Bảng 1.** Công cụ nghiên cứu

	Câu hỏi nghiên cứu	Công cụ
1	Nhận thức của giảng viên về hai thang đo đánh giá kỹ năng nói tiếng Anh, thang đo tổng quát và cụ thể.	Bảng câu hỏi Phỏng vấn
2	Nhận thức của giảng viên được phản ánh như thế nào trong thực hành sử dụng thang đo đánh giá tổng quát và cụ thể trong đánh giá kỹ năng nói?	Phỏng vấn

Nghiên cứu này là một nghiên cứu mô tả. Cả dữ liệu định lượng và định tính đều được thu thập và phân tích để hiểu đầy đủ vấn đề nghiên cứu và củng cố kết quả (Gay, Mills & Airasian, 2011). Trong nghiên cứu này, dữ liệu định lượng được thu thập thông qua bảng câu hỏi để giải quyết câu hỏi nghiên cứu đầu tiên, tức là để tìm hiểu nhận thức của giảng viên về sử dụng thang đo tổng quát và cụ thể trong việc đánh giá khả năng nói. Bên cạnh đó, để hiểu sâu hơn về nhận thức của giảng viên các trường đại học về thang đo tổng quát và cụ thể trong việc đánh giá khả năng nói, dữ liệu định tính đã được sử dụng qua các cuộc phỏng vấn. Ngoài ra, để tìm hiểu xem nhận thức của giảng viên được phản ánh như thế nào trong thực hành được báo cáo của họ về việc sử dụng các thang đo, dữ liệu định tính cũng được thu thập thông qua cuộc phỏng vấn.

### 4.4 Bảng câu hỏi

Trọng tâm của bảng câu hỏi là về nhận thức của giảng viên về thang đo đánh giá tổng quát và cụ thể trong đánh giá kỹ năng nói. Bảng câu hỏi bao gồm khoảng 49 câu hỏi với 4 nhóm.

**Bảng 2.** Mô tả nhóm câu hỏi

<b>Nhóm câu hỏi</b>	<b>Nội dung hỏi</b>
Thông tin ứng viên	Tên, tuổi, trình độ chuyên môn, kinh nghiệm giảng dạy và tham dự các khóa đào tạo / hội thảo giảng dạy.
Tầm quan trọng của việc sử dụng thang đo trong đánh giá kỹ năng nói	Sự cải thiện năng lực giao tiếp, sự hướng dẫn giảng dạy, sự tự định hướng của người học, tăng cường động cơ học tập và giảm đi sự thiên vị trong số những người đánh giá, sự hiểu biết về yêu cầu của chương trình giảng dạy, sự cần thiết của việc sử dụng các thang đo.
Thang đo tổng quát	Định nghĩa, các tính chất đặc trưng của thang đo tổng quát.
Thang đo cụ thể	Định nghĩa, các tính chất đặc trưng của thang đo cụ thể.

Nhóm thứ nhất gồm 7 câu hỏi để hỏi thông tin cơ bản của người tham gia bao gồm tên, tuổi, trình độ chuyên môn, kinh nghiệm giảng dạy và tham dự các khóa đào tạo / hội thảo giảng dạy. Nhóm thứ hai bao gồm 7 câu hỏi để điều tra nhận thức của giảng viên về tầm quan trọng của việc sử dụng thang đo đánh giá trong đánh giá kỹ năng nói. Đối với nhóm tiếp theo gồm 17 câu hỏi về nhận thức của giảng viên về các thang đo đánh giá tổng quát trong việc đánh giá khả năng nói. Nhóm cuối cùng gồm 18 câu hỏi để khám phá nhận thức của giảng viên về thang đo cụ thể. Những câu hỏi này được thiết lập theo thang đo năm bậc của Likert với năm điểm (rất không đồng ý, không đồng ý, trung lập, đồng ý, rất đồng ý).

#### 4.5 Phỏng vấn

Do quá trình đánh giá kỹ năng nói xuất hiện trong đầu của người đánh giá nên không thể nhìn thấy họ thực hành nó như thế nào (Fulcher & Davidson, 2007). Cách dễ tiếp cận nhất để xem thực hành của họ là thực hiện cuộc phỏng vấn để xem việc họ sử dụng thang đo tổng quát và cụ thể trong đánh giá kỹ năng nói như thế nào qua báo cáo của họ. 9 giảng viên được chọn ngẫu nhiên để phỏng vấn. Câu hỏi phỏng vấn tập trung vào 2 điểm: lý do tại sao giảng viên ưa thích thang đo đánh giá tổng quát hay cụ thể và nhận thức của họ được phản ánh như thế nào qua việc sử dụng thang đo đánh giá tổng quát và cụ thể trong đánh giá kỹ năng nói tiếng Anh.

Dựa trên các câu hỏi từ bảng câu hỏi, 2 câu hỏi chính được thiết kế để tìm hiểu lý do tại sao giảng viên lại ưa thích thang đo đánh giá tổng quát và cụ thể. 2 câu hỏi phỏng vấn khác đã được sử dụng để tìm hiểu xem nhận thức của giảng viên được phản ánh như thế nào trong thực hành được báo cáo của họ về việc sử dụng các thang đo đánh giá tổng quát và cụ thể trong đánh giá kỹ năng nói. Mỗi câu hỏi nhằm khám phá quy trình mà giảng viên thực hành thang đo đánh giá tổng quát và cụ thể. Trong quy trình này, các giảng viên đã báo cáo từng bước chi tiết của việc sử dụng thang đo đánh giá tổng quát và cụ thể. Cuối cùng, tất cả các câu trả lời của những người tham gia được ghi chép và phân tích để tìm ra nhận thức của họ thông qua thực hành được phản ánh.

#### 4.6 Phân tích số liệu

Câu trả lời của những người tham gia cho bảng câu hỏi được mã hóa thành số. Đầu tiên, Scale Test được chạy để kiểm tra độ tin cậy của bảng câu hỏi. Tiếp theo, thống kê mô tả (Descriptive Statistics) được tính toán để quan sát điểm trung bình (Mean), điểm tối đa

(Max), điểm tối thiểu (Min) và độ lệch chuẩn (Sd) của cả nhóm. Cuối cùng, One-sample T-test được thực hiện để đánh giá xem liệu điểm trung bình của nhận thức của giảng viên về các thang đo đánh giá tổng quát và cụ thể trong việc đánh giá khả năng nói có khác biệt về mặt thống kê với giá trị kiểm định hay không.

Các câu trả lời của những người tham gia từ các cuộc phỏng vấn đã được chép lại, đọc và phân tích để có được những hiểu biết sâu sắc về đề tài nghiên cứu này. Bên cạnh đó, các cuộc phỏng vấn cho thấy việc thực hành phản ánh của giảng viên về việc sử dụng thang đo tổng quát và phân tích trong đánh giá bài kỹ năng nói. Giai đoạn cuối cùng của quá trình này là giải thích dữ liệu thu thập được.

## 5. KẾT QUẢ

Sau khi các câu trả lời của ứng viên từ bảng câu hỏi được nhập liệu và mã hóa, độ tin cậy của bảng câu hỏi được kiểm tra. Kết quả độ tin cậy (reliability coefficient) của cả bảng câu hỏi  $\alpha = .796$ . Hệ số tin cậy ( $\alpha = .796$ ) có thể chấp nhận. Điều cho thấy dữ liệu được sử dụng trong nghiên cứu này là đáng tin.

### 5.1 Nhận thức chung của giảng viên về thang đo tổng quát và thang đo cụ thể trong đánh giá kỹ năng nói tiếng Anh

**Bảng 3.** Nhận thức chung của giảng viên về thang đo đánh giá tổng quát và thang đo cụ thể khi sử dụng cho việc đánh giá kỹ năng nói tiếng Anh.

Biến	N	Min.	Max.	M	SD
Nhận thức của giảng viên về thang đo đánh giá tổng quát và thang đo cụ thể	40	3.47	4.58	3.99	.249

**Bảng 4.** Kiểm định T-test về nhận thức chung của giảng viên về thang đo đánh giá tổng quát và thang đo cụ thể khi sử dụng cho việc đánh giá kỹ năng nói tiếng Anh.

	Giá trị kiểm định = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval	
					Lower	Upper
Mean tổng	-,114	39	,910	-,00449	-,0840	,0750

Từ **Bảng 3** có thể thấy rằng tổng điểm trung bình về nhận thức của người tham gia ( $M = 3,99$ ) nằm trong thang điểm 3 (trung bình) và 4 (cao) trong thang đo năm bậc Likert. Sau đó, One Sample T-Test đã được chạy để kiểm tra xem liệu mức độ nhận thức của người tham gia về thang đo đánh giá tổng quát và cụ thể trong việc đánh giá khả năng nói ( $M = 3,99$ ,  $SD = 0,249$ ) có khác biệt về mặt thống kê với giá trị kiểm định là 4,0. Kết quả (**Bảng 4**) cho thấy không có sự khác biệt đáng kể giữa giá trị trung bình của mẫu ( $M = 3,99$ ,  $SD = 0,249$ ) và giá trị thử nghiệm 4,0 ( $t = -.114$ ,  $df = 39$ ,  $p = .910$ ). Kết quả cho thấy rằng nhận thức của giảng viên được điều tra là cao.

## 5.2 Nhận thức của giảng viên về tầm quan trọng của việc sử dụng thang đo trong đánh giá kỹ năng nói

**Bảng 5.** Nhận thức của giảng viên về tầm quan trọng của việc sử dụng thang đo trong đánh giá kỹ năng nói

Biến	N	Min.	Max.	M	SD
Nhận thức của giảng viên về tầm quan trọng của việc sử dụng thang đo trong đánh giá kỹ năng nói	40	3.29	5.00	4.14	.394

**Bảng 6.** One-sample test về nhận thức của giảng viên về tầm quan trọng của việc sử dụng thang đo trong đánh giá kỹ năng nói (Giá trị kiểm định = 4)

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval	
					Lower	Upper
Mean nhóm 1	2,181	39	,035	,13571	,0099	,2616

Như được trình bày trong **bảng 5**, có thể thấy rõ rằng những người tham gia có nhận thức tích cực về các khía cạnh được điều tra. Các giảng viên nhận thức về tầm quan trọng của việc sử dụng thang đo trong việc đánh giá khả năng nói ( $M = 4,14$ ,  $SD = 0,394$ ) là nhóm có nhận thức cao nhất.

One-sample T-test đã được chạy để kiểm tra xem mức độ nhận thức của những người tham gia về việc sử dụng các điểm đánh giá trong việc đánh giá khả năng nói ( $M = 4,14$ ,  $SD = 0,394$ ) có khác biệt về mặt thống kê so với giá trị kiểm tra 4,0 hay không. Giá trị trung bình mẫu ( $M = 4,14$ ,  $SD = 0,394$ ) khác biệt đáng kể so với giá trị thử nghiệm 4,0 ( $t = 2,18$ ,  $df = 39$ ,  $p = 0,035$ ). Kết quả (**bảng 6**) cho thấy rằng nhận thức của giáo viên được khảo sát về tầm quan trọng của việc sử dụng phiếu đánh giá trong việc đánh giá khả năng nói là tương đối cao.

## 5.3 Nhận thức của giảng viên về thang đo đánh giá tổng quát trong việc đánh giá kỹ năng nói tiếng Anh.

**Bảng 7.** Nhận thức của giảng viên về thang đo đánh giá tổng quát trong đánh giá kỹ năng nói tiếng Anh.

Biến	N	Min.	Max.	M	SD
Nhận thức của giảng viên về thang đo tổng quát trong đánh giá kỹ năng nói tiếng Anh	40	3.35	4.29	3.84	.264

**Bảng 8.** One-sample T-test về nhận thức của giảng viên về thang đo đánh giá tổng quát trong việc đánh giá kỹ năng nói tiếng Anh (Giá trị kiểm định = 3)

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval	
					Lower	Upper
Mean nhóm 2	20,137	39	,000	,83971	,7554	,9241

**Bảng 9.** One-Sample Test về nhận thức của giảng viên về tầm quan trọng của việc sử dụng thang đo trong đánh giá kỹ năng nói (Giá trị kiểm định = 4)

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval	
					Lower	Upper
Mean nhóm 2	-3,844	39	,000	-,16029	-,2446	-,0759

One-sample T-test được chạy để kiểm tra xem liệu mức độ nhận thức của người tham gia về các điểm đánh giá tổng quát trong việc đánh giá khả năng nói (**bảng 7**) có khác biệt về mặt thống kê với giá trị thử nghiệm là 3,0, một giá trị trung bình. Giá trị trung bình của mẫu khác biệt đáng kể so với giá trị thử nghiệm 3.0 ( $t = 20,14$ ,  $df = 39$ ,  $p = 0,00$ ) (**bảng 8**).

One-sample T-test được chạy để kiểm tra xem liệu mức độ nhận thức của người tham gia về các điểm đánh giá tổng quát trong việc đánh giá khả năng nói ( $M = 3,84$ ,  $SD = 0,264$ ) (**bảng 7**) có khác biệt về mặt thống kê với giá trị thử nghiệm là 4,0 hay không, một giá trị trung bình cao. Giá trị trung bình của mẫu ( $M = 3,84$ ,  $SD = 0,264$ ) khác biệt đáng kể so với giá trị thử nghiệm 4,0 ( $t = -3,85$ ,  $df = 39$ ,  $p = 0,00$ ) (**bảng 9**). Kết quả cho thấy nhận thức của giáo viên được điều tra về các điểm đánh giá tổng quát trong việc đánh giá khả năng nói chỉ trên mức trung bình.

#### 5.4 Nhận thức của giảng viên về thang đo cụ thể trong đánh giá kỹ năng nói tiếng Anh.

**Bảng 10.** Nhận thức của giảng viên về thang đo cụ thể trong việc đánh giá kỹ năng nói tiếng Anh.

Biến	N	Min.	Max.	M	SD
Nhận thức của giảng viên về thang đo cụ thể trong việc đánh giá kỹ năng nói tiếng Anh.	40	3.44	4.67	4.01	.324

**Bảng 11.** One-sample test về nhận thức của giảng viên về thang đo cụ thể trong việc đánh giá kỹ năng nói tiếng Anh (Giá trị kiểm định = 4)

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval	
					Lower	Upper
Mean nhóm 3	,217	39	,829	,01111	-,0923	,1146

One-sample T-test được chạy để kiểm tra xem mức độ nhận thức của những người tham gia đối với thang đo cụ thể trong việc đánh giá khả năng nói (**bảng 10**) có khác biệt về mặt thống kê với giá trị kiểm tra 4.0 hay không, một mức trung bình cao. Kết quả chỉ ra rằng không có sự khác biệt giữa giá trị trung bình của mẫu ( $M = 4.01$ ,  $SD = .324$ ) và giá trị thử nghiệm 4.0 ( $t = .217$ ,  $df = 39$ ,  $p = .829$ ) (**bảng 11**). Điều này đã chỉ ra rằng nhận thức của giáo viên về thang đo cụ thể trong việc đánh giá khả năng nói là cao.

#### 5.5 Giải thích của ứng viên khi sử dụng thang đo tổng quát và thang đo cụ thể trong việc đánh giá kỹ năng nói tiếng Anh.

Trong số 9 ứng viên được phỏng vấn, 4 ứng viên thích dùng kết hợp cả hai thang đo vì họ nghĩ điều này sẽ làm cho kết quả mang tính chính xác hơn và nhất quán giữa các giám khảo chấm bài, 2 ứng viên thích sử dụng thang đo tổng quát hơn vì họ cho rằng thang đo này giúp họ tiết kiệm nhiều thời gian, dễ sử dụng và linh hoạt hơn, 3 ứng viên lại thích sử dụng thang đo cụ thể hơn và họ đồng tình rằng thang đo này chính xác, đáng tin, nhất quán và hữu ích.

## **5.6 Việc sử dụng thang đo tổng quát và thang đo cụ thể trong thực tế**

### **5.6.1 Loại thang đo các ứng viên thường sử dụng**

Trong số 9 giảng viên được phỏng vấn, 4 giảng viên thường sử dụng thang đo cụ thể vì đó là quy định của trường. Tất cả các giáo viên này đều đồng ý 4 thang đo này đáng tin cậy.

3 giảng viên được phỏng vấn lại cho rằng sử dụng thang đo tổng quát giúp họ tiết kiệm nhiều thời gian hơn. Họ cảm thấy thoải mái hơn và tự do hơn.

Các giảng viên còn lại cho rằng việc kết hợp hai thang đo mang lại độ tin cậy cao hơn và tính minh bạch được giữa các giám khảo chấm bài cao hơn.

### **5.6.2 Thực tế sử dụng thang đo tổng quát trong việc đánh giá kỹ năng nói.**

Khi sử dụng thang đo này, các giảng viên chuẩn bị một bảng tiêu chí đánh giá mang tính chung chung, khái quát. Trong khi đánh giá, họ lắng nghe và quan sát người nói. Sau đó, dựa trên ấn tượng của họ về sự thể hiện của người nói để mà cho điểm. Quá trình cho điểm diễn ra trong suy nghĩ của giáo viên một cách nhanh chóng.

### **5.6.3 Thực tế sử dụng thang đo cụ thể trong việc đánh giá kỹ năng nói.**

Đầu tiên, một bảng tiêu chí đánh giá được thiết lập. Các tiêu chí này bao gồm ngữ pháp, từ vựng, ngữ âm, ngữ điệu, sự thành thạo, tính chính xác và nội dung. Tất cả các tiêu chí này được phân chia ở các cấp độ chi tiết và được chia điểm cụ thể để đánh giá người nói. Kế đến, giám khảo sẽ nghe, quan sát và ghi điểm cho từng mức độ ở mỗi tiêu chí mà thí sinh đạt được. Điểm số cuối cùng sẽ là điểm tổng của từng tiêu chí mà thí sinh đó đạt được. Thực tế cho thấy, sử dụng thang đo này mang lại độ tin cậy cao hơn và tránh sự không nhất quán trong khi cho điểm số giữa các giám khảo.

### **5.6.4 Thực tế sử dụng kết hợp hai thang đo tổng quát và cụ thể trong việc đánh giá kỹ năng nói.**

Khi đánh giá kỹ năng nói, đầu tiên các giám khảo sẽ sử dụng thang đo tổng quát để đánh giá toàn bộ phần thể hiện của thí sinh dựa trên ấn tượng của mình. Sau đó, các tiêu chí đánh giá của thang đo cụ thể được sử dụng để phân tích chi tiết hơn về phần thể hiện của thí sinh. Điều này làm tăng độ chính xác của việc đánh giá. Cuối cùng các giám khảo sẽ cộng lại tất cả các tiêu chí để cho ra điểm số cuối cùng.

## **6. THẢO LUẬN**

Kết quả của nghiên cứu cho thấy phương pháp đánh giá theo thang đo cụ thể được cho là chính xác và đáng tin cậy hơn phương pháp đánh giá theo thang đo tổng quát. Kết quả này tương ứng với kết quả nghiên cứu được thực hiện bởi Bakaoui (2010). Ông so sánh các giảng viên chưa có kinh nghiệm và có kinh nghiệm về nhận thức của họ về thang đo tổng quát và cụ thể. Kết quả chỉ ra rằng cả giảng viên chưa có kinh nghiệm và có kinh nghiệm có xu hướng thích thang đo đánh giá cụ thể. Tương tự, Bahmani (2014) đã điều tra các kết quả ngược của việc chấm điểm kỹ năng nói dựa trên thang đo cụ thể và thang đo tổng quát của người học tiếng Anh như ngôn ngữ thứ hai ở Iran. Kết quả chỉ ra rằng các cách đánh giá theo thang đo cụ thể có tác dụng tích cực hơn trong việc cải thiện hiệu suất nói của người học so với cách đánh giá theo thang đo tổng thể. Tương ứng, Metruk (2018) đã so sánh các cách đánh giá theo thang đo tổng quát và cụ thể trong đánh giá kỹ năng nói. Kết quả cho thấy phương pháp đánh giá theo thang đo cụ thể đã được chứng minh là chính xác và đáng tin cậy hơn so với phương pháp đánh giá theo thang đo tổng quát.



Phân tích từ dữ liệu định lượng và định tính cho thấy nhận thức của giảng viên về phương pháp đánh giá theo thang đo tổng quát thấp hơn nhận thức theo thang đo cụ thể. Các ứng viên cho rằng thang đo tổng quát là ít chính xác và đáng tin cậy. Cũng chính vì thế mà muốn kiểm tra độ chính xác của việc đánh giá thường các giảng viên lại sử dụng thang đo đánh giá cụ thể. Ngược lại, Iwashita và Grove (2003) đã so sánh các thang đo tổng quát và cụ thể trong việc đánh giá khả năng nói. Kết quả cho thấy rằng thang đo tổng quát có nhiều lợi thế hơn so với thang đo cụ thể. Maha (2017) đã rút ra một kết luận từ nghiên cứu của mình rằng cách đánh giá tổng quát có độ tin cậy tương đối cao và mức độ nhất quán giữa các giám khảo đánh giá cao. Tuy nhiên, trong nghiên cứu này, phân tích từ dữ liệu cho thấy rằng thang đo tổng quát là ít chính xác hơn và ít đáng tin cậy hơn thang đo cụ thể. Lý do là quá trình đánh giá không rõ ràng và dễ hiểu. Hơn nữa, hầu hết các đánh giá đều dựa trên ấn tượng của giảng viên về thể hiện của người nói. Quá trình đánh giá và ghi điểm mang tính chủ quan.

## 7. KIẾN NGHỊ

Từ những phát hiện trên, một số kiến nghị được đề xuất nhằm cho giảng viên và các nhà quản lý giáo dục.

Để đảm bảo độ tin cậy của việc đánh giá kỹ năng nói, giảng viên nên chọn thang đo đánh giá cụ thể. Từ việc phân tích dữ liệu định tính, rõ ràng thang đo đánh giá cụ thể có lợi cho giảng viên trong việc tăng mức độ chính xác và độ tin cậy. Một số giảng viên sử dụng thang đo tổng quát đầu tiên. Nếu học sinh cần làm rõ các kết quả, giảng viên sẽ sử dụng thang đo phân tích sau đó. Nếu giảng viên cần phân tích điểm mạnh hay điểm yếu của sinh viên và để cho những phản hồi tích cực cho sinh viên, họ cũng sẽ sử dụng thang đo cụ thể. Mặc dù thang đo tổng quát là sự lựa chọn đầu tiên của giảng viên trong việc đánh giá kỹ năng nói, họ cũng sử dụng các thang đo cụ thể như là lựa chọn tối ưu của họ để đảm bảo tính chính xác và nhất quán của các người đánh giá.

Thang đo đánh giá cụ thể giúp giảng viên nhìn thấy điểm mạnh và điểm yếu của người học của mình một cách chi tiết hơn. Do đó, họ có thể áp dụng các phương pháp giảng dạy phù hợp để giúp học sinh của mình tăng cường điểm mạnh và khắc phục các điểm yếu.

Về mặt thời gian, thang đo tổng quát tiết kiệm thời gian hơn. Trong khi đó, thang đo cụ thể tốt hơn trong việc cho những nhận xét tích cực. Nó phụ thuộc vào loại bài kiểm tra mà giảng viên sẽ chọn. Kiểm tra tổng kết nên sử dụng thang đo tổng quát. Trong khi đó, kiểm tra quá trình nên sử dụng thang đo cụ thể.

Thang đo cụ thể được sử dụng khi kiểm tra ở cuối một đơn vị hoặc bài học cụ thể sẽ hữu ích. Trong khi đó, thang đo tổng quát thì hữu ích vào cuối khóa học. Do đó, sau khi đánh giá sử dụng thang đo cụ thể được áp dụng vào cuối mỗi bài học học sinh trở lại lớp học và nhận được phản hồi và kết quả. Những phản hồi này có thể giúp người học nắm bắt được điểm mạnh và điểm yếu của họ. Đây là một động lực để họ cải thiện kỹ năng nói cho các lần đánh giá tiếp theo. Trong khi đó, thang đo tổng quát hữu ích vào cuối khóa học. Sinh viên không có cơ hội quay trở lại lớp học và nhận phản hồi. Hơn nữa, không có đánh giá tiếp theo. Để tiết kiệm thời gian, giảng viên thường sẽ chọn thang đo đánh giá tổng quát nếu không có thắc mắc về điểm số cuối cùng từ sinh viên.

Cách đánh giá nên được công khai cho các sinh viên khi bắt đầu khóa học để họ có thể biết những yêu cầu của chương trình học. Khi họ biết mục tiêu của chương trình, họ có thể có thêm động lực và tự định hướng trong học tập nhờ những yêu cầu rõ ràng từ các tiêu chí của thang điểm đánh giá.

Điều cần thiết là phải thiết lập các tiêu chí rõ ràng trước khi đánh giá. Điều này giúp giảm sự thiên vị và sự không nhất quán giữa những người đánh giá. Đồng thời, nó cũng giúp nâng cao độ tin cậy của việc đánh giá.

### TÀI LIỆU THAM KHẢO

- [1] Alderson, J. C., Clapham, C., & Wall, D. 1995. *Language test construction and evaluation*. Cambridge, UK: Cambridge University Press
- [2] Andrade, H. 2000. Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18
- [3] Arter, J. 2000. Rubrics, scoring guides, and performance criteria: Classroom tools for assessing and improving student learning. Paper presented at the annual conference of the American Educational Research Association, New Orleans.
- [4] Barkaoui, K. 2010. Variability in ESL essay rating processes: The role of rating scale and rater experience. *Language Assessment Quarterly*, 7(1), 54-74.
- [5] Black, P., & William, D. 1998. *Assessment in education: Principles, policy and practice*. London: Routledge.
- [6] Brown, H. D. 2004. *Language assessment: Principles and classroom practices*. NY, USA: Pearson Education.
- [7] Chaney, A.L., & Burk, T.L. 1998. *Teaching oral communication in grades K-8*. Boston: Allyn & Bacon.
- [8] Cohen, L., Manion, L., & Morrison, K. 2000. *Research methods in education (5th ed.)*. London: Routledge/Falmer.
- [9] Creswell, J. W. 2003. *Research design: qualitative, quantitative and mixed methods approaches (2nd ed.)*. London: Sage Publications Ltd.
- [10] Edmunds, J. 2006. *How to assess student performance in history*. Greensboro: The SERVE center at the university of North Carolina.
- [11] Fraenkel, J. R., E. Wallen & H. H. Huyn. 1993. *How to design and evaluate research in education*. New York: McGraw-Hill
- [12] Fulcher, G. 2003. *Testing second language speaking*. Harlow, UK: Longman
- [13] Fulcher, G. 2010. *Practical language testing*. London: Hodder Education.
- [14] Fulcher, G., & Davidson, F. 2007. *Language testing and assessment. An advance resource book*. Oxford: Routledge
- [15] Gay, L., Mills, G., & Airasian, P.(2011). *Educational research (10th ed.): Competencies for analysis and applications*. New Jersey: Pearson Education, Inc.
- [16] Gerard. 2000. *How to teach pronunciation*. England: Pearson longman education

- 
- [17] Hornby, A.S. 1995. Oxford advanced learners' dictionary of current language. Oxford: Oxford university press.
- [18] Hughes, A. 2003. Testing for language teachers (2nd ed.). Cambridge, UK: Cambridge university press
- [19] Huot, B. 1990. Reliability, validity, and holistic scoring: What we know and what we need to know. *College Composition and Communication*, 41(2), 201-213.
- [20] Hymes, D. 1974. Foundations in sociolinguistics: An ethnographic approach. Philadelphia: University of Pennsylvania Press.
- [21] Johnson, K. E. 1994. The emerging beliefs and instructional practices of preservice English as a second language teachers. *Teaching and teacher education*, 10(4), 439-452.
- [22] Knoch, U. 2009. Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275-304.
- [23] Larsen-Freeman, D. 2000. Techniques and principles in language teaching. Oxford: Oxford University Press.
- [24] Lumley, T. 2005. Assessing second language writing: The raters' perspective. Frankfurt: Peter Lang.
- [25] Luoma, S. 2004. Assessing speaking. Cambridge, UK: Cambridge University Press.
- [26] Luu, T.T. 2012. Teaching and assessing speaking performance through analytic scoring approach. *Theory and Practice in Language Studies*, 2(4), 673-679. doi: 10.4304/tpls.2.4.673-679
- [27] Maha, O. 2017. A comparison between holistic and analytic assessment of speaking. *Journal of Language Teaching and Research*, 8(4), 679-690. doi: 10.17507/jltr.0804.06.
- [28] McNamara, T. 2000. Language testing. Oxford, UK: Oxford University Press.
- [29] Mead, N.A., & Rubin, D.L. 1985. Assessing listening and speaking skill. Eric. Retrieved from <https://eric.ed.gov/?id=ED263626>
- [30] Mertler, C. A. 2001. Designing scoring rubrics for your Classroom. *Practical Assessment, Research & Evaluation*, 7(25), 1-8
- [31] Metruk, R. 2018. Comparing holistic and analytic ways of scoring in the assessment of speaking skill. *Journal of Teaching English for Specific and Academic Purposes*, 6(1), 179-189. doi: 10.22190/JTESAP1801179M
- [32] Phuong, H. Y. & Phan, T.X. 2017. Using analytic rubric for speaking self-assessment: EFL students' perceptions and challenges. *Research and Method in Education*, 7(3), 34-39. Retrieved from <http://www.iosrjournals.org/>
- [33] Reddy, Y.M. 2007. Effects of rubrics on enhancement of student learning. *Educatejournal*, 7(1), 3-17. Retrieved from <http://www.educatejournal.org/index.php/educate/article/viewFile/117/148>
- [34] Richards, J. C. & Lockhart, C. 1994. Reflective teaching in second language classrooms. England: Cambridge University Press.
- [35] Rhodes, L.K., & Sally, N.M. 1992. Anecdotal records: a powerful tool for ongoing literacy assessment. *Jstor*, 45(7), 502-509. Retrieved from <http://www.jstor.org/stable/20200910>
-

- [36] Richards, J. C. 1983. Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17(2), 219-240.
- [37] Sharon, K. 2008. Using rubrics to measure and enhance student performance. Nova. Retrieved from [www.nvcc.edu](http://www.nvcc.edu).
- [38] Stiggins, R. J. 2001. *Student-involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Merrill/Prentice-Hall.
- [39] Syakur. 1987. *Language testing and evaluation*. Surakarta: Sebelas Maret University Press
- [40] Thomas, K., & Vincent, G. 2001. *Using assessment to improve the quality of education*. Paris: UNESCO.
- [41] Ur, P. 1996. *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.
- [42] Weigle, S. C. 2002. *Assessing writing*. Cambridge: Cambridge University Press.
- [43] Weir, C. J. 2005. *Language testing and validation - An evidence based approach*. New York: Palgrave Macmillan.
- [44] Wilkins, D. 1972. *Linguistics in language teaching*. London: Arnold.