

# Nghiên cứu quan điểm của giáo viên về việc sử dụng trí tuệ nhân tạo trong đánh giá bài viết IELTS

Hồ Nguyễn Thái Oanh<sup>1</sup>, Văn Phạm Kỳ An\*<sup>2</sup>

<sup>1</sup> Email: oanhho.master@gmail.com

Trường Đại học Văn Lang  
45 Nguyễn Khắc Nhu, Quận 1,  
Thành phố Hồ Chí Minh, Việt Nam

\* Tác giả liên hệ

<sup>2</sup> Email: anvan@alumni.upenn.edu  
Trường Đại học Pennsylvania  
3700 đường Walnut, Philadelphia,  
Pennsylvania, Hoa Kỳ

**TÓM TẮT:** Trong bối cảnh ngày càng nhiều trường đại học sử dụng chứng chỉ tiếng Anh quốc tế IELTS trong quy trình tuyển sinh, nhu cầu học IELTS ngày càng gia tăng. Do đó, các giáo viên phải dành nhiều thời gian hơn để chấm bài viết IELTS nhằm đáp ứng nhu cầu ngày càng cao của học sinh. Trước sự phổ biến của ChatGPT, một mô hình trí tuệ nhân tạo (AI) ra mắt vào tháng 11 năm 2022, nhiều giáo viên bắt đầu sử dụng ChatGPT để chấm điểm bài viết IELTS nhằm tiết kiệm thời gian và nâng cao chất lượng đánh giá. Tuy nhiên, không phải tất cả giáo viên đều ứng dụng công nghệ mới này một cách hiệu quả. Bài viết tìm hiểu quan điểm của giáo viên IELTS về các thách thức và chiến lược tối ưu để tích hợp AI trong đánh giá bài viết IELTS. Nghiên cứu áp dụng phương pháp hỗn hợp kết hợp bảng câu hỏi khảo sát định lượng và phỏng vấn định tính để thu thập dữ liệu. Kết quả cho thấy thách thức chính liên quan đến cách tạo lập câu lệnh (prompt). Bên cạnh đó, nghiên cứu đề xuất cách tạo câu lệnh hiệu quả và xây dựng cộng đồng giáo viên để trao đổi kinh nghiệm sử dụng AI. Những phát hiện mới trong nghiên cứu sẽ giúp các nhà giáo dục tối ưu hoá việc sử dụng AI trong đánh giá kỹ năng viết.

**TỪ KHÓA:** Trí tuệ nhân tạo (AI), ChatGPT, IELTS, đánh giá năng lực ngôn ngữ, Công nghệ giáo dục.

→ Nhận bài 09/5/2024 → Nhận bài đã chỉnh sửa 28/6/2024 → Duyệt đăng 20/8/2024.

DOI: <https://doi.org/10.15625/2615-8957/12410805>

## 1. Đặt vấn đề

IELTS (International English Language Testing System) là bài thi quốc tế nhằm kiểm tra năng lực tiếng Anh của thí sinh ở bốn kỹ năng: Nghe, Nói, Đọc và Viết. Theo Hội đồng Anh (British Council), IELTS là kì thi năng lực tiếng Anh được công nhận rộng rãi nhất thế giới cho giáo dục đại học và di cư quốc tế với hơn bốn triệu bài thi hằng năm [1]. Ở Việt Nam, xu hướng các trường đại học sử dụng chứng chỉ IELTS trong việc tuyển sinh mỗi năm ngày càng tăng. Tính đến tháng 3 năm 2024, có hơn 50 trường đại học uy tín trên cả nước đánh giá ứng viên tiềm năng dựa trên điểm IELTS kết hợp với thành tích học thuật của học sinh trong suốt thời gian học trung học [2]. Chính vì thế, nhu cầu học IELTS đang tăng mạnh tại Việt Nam. Điều này dẫn đến việc các giáo viên phải nỗ lực hơn để đáp ứng nhu cầu tăng cao của học sinh trong khi vẫn đảm bảo đánh giá bài viết một cách chất lượng. Trong bối cảnh hiện nay, việc chấm điểm thủ công tốn nhiều thời gian và đòi hỏi nhiều công sức hơn trước. Theo một nghiên cứu năm 2020, giáo viên dạy IELTS ở Việt Nam cần dành khoảng 15 phút để đánh giá một bài viết luận IELTS [3]. Điều này có nghĩa là cần 5 giờ để đánh giá 20 bài viết hoàn thành bởi 10 học sinh. Ngoài việc đánh giá, giáo viên IELTS phải chịu trách nhiệm nhiều công tác hành chính khác. Vì thế, giáo viên IELTS đều cảm thấy

căng thẳng. Hơn nữa, để duy trì tính nhất quán và độ chính xác trong việc đánh giá bài viết của học sinh, giáo viên phải trải qua một quá trình đào tạo kỹ lưỡng. Mặc dù có một phiếu tiêu chí đánh giá được cung cấp công khai bởi ban tổ chức cuộc thi IELTS nhưng giáo viên vẫn cần được đào tạo chuyên sâu hơn về quy trình đánh giá và cách đưa ra góp ý mang tính xây dựng để giúp học sinh cải thiện kỹ năng viết [4]. Do đó, giáo viên và người đánh giá cần phải nỗ lực hơn nhiều để đảm bảo rằng học sinh IELTS nhận được góp ý và hướng dẫn có chiều sâu. Vì vậy, các nhà giáo dục, lãnh đạo trường học và giáo viên cần xem xét những cách thức thay thế để đơn giản hóa quá trình đánh giá mà vẫn duy trì hiệu quả và chất lượng đánh giá cao.

Một số phương pháp mới để đánh giá kỹ năng viết đang phát triển mạnh mẽ và một trong số đó là sử dụng AI. AI sử dụng các phương pháp thông minh như xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP) và học từ dữ liệu để cho điểm và đưa ra góp ý một cách tự động. Do đó, nghiên cứu của Chen, Chen, & Lin (2020) đưa ra rằng, AI cho phép giáo viên có nhiều thời gian hơn để nói chuyện và làm việc với học sinh vì nó có thể tự đánh giá bài làm. Họ đã chứng minh AI là một công cụ hỗ trợ đắc lực cho giáo viên, giúp tối ưu hóa quy trình đánh giá và giải phóng thời gian giúp giáo viên cá nhân hoá việc giảng dạy của mình cho từng

học viên [5]. Hơn nữa, cộng đồng giáo dục đang nghiên cứu tiềm năng của các trợ lý ảo được cung cấp bởi AI trong việc đánh giá bài viết của học sinh. Cleveland (2023) chia sẻ kinh nghiệm của họ trong việc sử dụng ChatGPT để đánh giá kỹ năng viết của học sinh [6]. Bằng những phép thử với nhiều câu lệnh khác nhau yêu cầu ChatGPT đưa ra đánh giá bài viết của học sinh, họ đã cho thấy tiềm năng của các trợ lý ảo AI trong việc đánh giá các bài viết một cách hiệu quả và giúp tiết kiệm thời gian cho giáo viên nếu người dùng biết “huấn luyện” AI bằng những câu lệnh rõ ràng.

Nhờ các tiến bộ công nghệ này, nhiều tổ chức giáo dục lớn đã triển khai AI trong quá trình đánh giá của họ như Khan Academy và Tổ chức Kiểm tra Giáo dục (ETS) [7], [8]. Trong khi đó, vẫn chưa nhiều giáo viên ở Việt Nam sử dụng AI trước năm 2022. Từ tháng 11 năm 2022, sự ra đời của ChatGPT, một mô hình ngôn ngữ lớn dễ sử dụng được điều khiển bởi AI đã giúp nhiều giáo viên tiếng Anh ở Việt Nam dễ dàng bắt đầu sử dụng công nghệ này. Tuy nhiên, không phải tất cả giáo viên đều sử dụng ChatGPT hiệu quả. Do đó, bài viết này nhằm mục đích nghiên cứu các thách thức, chiến lược mà các giáo viên IELTS đã sử dụng để đối phó với những thách thức và các đề xuất của họ về việc sử dụng công cụ AI trong việc đánh giá bài viết IELTS.

Trong khi các nghiên cứu hiện tại trên thế giới cung cấp một góc nhìn đa dạng về việc sử dụng AI trong đánh giá kỹ năng viết thì có một sự thiếu hụt đáng kinh ngạc về các nghiên cứu liên quan đến việc đánh giá kỹ năng viết sử dụng công nghệ AI tại Việt Nam. Nghiên cứu này của chúng tôi hi vọng góp phần vào kho tàng các nghiên cứu về sử dụng AI trong đánh giá kỹ năng viết và cung cấp gợi ý thiết thực cho giáo viên, hiệu trưởng, nhà phát triển chương trình học, chương trình đào tạo giáo viên và người thiết kế đề thi viết cũng như các nhà hoạch định chính sách giáo dục đang quan tâm đến việc đánh giá bài viết IELTS bằng cách sử dụng AI.

## 2. Nội dung nghiên cứu

### 2.1. Phương pháp nghiên cứu

Nghiên cứu này tìm hiểu quan điểm của giáo viên IELTS về việc áp dụng AI trong việc đánh giá các bài viết IELTS. Chúng tôi đặt mục tiêu trả lời ba câu hỏi nghiên cứu sau:

*Câu hỏi 1:* Những thách thức chính mà các giáo viên IELTS gặp phải khi tích hợp ChatGPT vào việc đánh giá bài viết IELTS?

*Câu hỏi 2:* Giáo viên IELTS đã làm thế nào để đối phó và vượt qua những thách thức này?

*Câu hỏi 3:* Các giáo viên IELTS đề xuất gì để áp dụng ChatGPT vào việc đánh giá bài viết IELTS một cách hiệu quả?

Để đạt được mục tiêu này, nghiên cứu áp dụng

phương pháp kết hợp thu thập dữ liệu định lượng bằng phiếu khảo sát và thu thập dữ liệu định tính thông qua các cuộc phỏng vấn. Những vấn đề đạo đức về tính hợp lệ và độ tin cậy ở mỗi giai đoạn nghiên cứu được xem xét chặt chẽ. Quy trình nghiên cứu bao gồm: 1) Tạo ra các công cụ thu thập dữ liệu; 2) Thử nghiệm mẫu; 3) Thu thập dữ liệu chính. Sau đó, chúng tôi thảo luận về phương pháp phân tích dữ liệu, tập trung vào việc sử dụng cả thống kê mô tả định lượng và phân tích định tính để rút ra những thông tin quan trọng từ dữ liệu đã thu thập. Quá trình thu thập dữ liệu chính kéo dài từ ngày 22 tháng 02 đến ngày 01 tháng 3 năm 2024.

#### 2.1.1 Phiếu khảo sát

Đầu tiên, chúng tôi tiến hành các cuộc phỏng vấn không chính thức với những giáo viên đã sử dụng AI để đánh giá các bài viết IELTS để thiết kế phiếu khảo sát một cách cẩn thận. Các mục đo lường theo thang điểm Likert được tích hợp vào phiếu khảo sát để người tham gia xếp hạng và thảo luận về các khía cạnh khác nhau của AI trong đánh giá bài viết IELTS. Các mục được xếp hạng từ 1 đến 5 với: 1/ Hoàn toàn không đồng ý với mệnh đề đã cho; 2/ Không đồng ý với mệnh đề đã cho; 3/ Trung lập; 4/ Đồng ý với mệnh đề đã cho; 5/ Hoàn toàn đồng ý với mệnh đề đã cho. Các mục tiêu chính của phiếu khảo sát là đo lường nhận thức và kiến thức của giáo viên về công nghệ AI, thái độ của họ đối với việc tích hợp AI vào đánh giá viết IELTS, quan điểm về ưu và nhược điểm của AI và sự sẵn lòng áp dụng và triển khai AI trong công tác đánh giá của họ. Ngoài ra, các câu hỏi về nhân khẩu học cũng được đưa vào phiếu khảo sát để thu thập thông tin về hình thức và kinh nghiệm giảng dạy của người tham gia, chẳng hạn như tuổi, giới tính, trình độ học vấn, bằng cấp giảng dạy, số năm kinh nghiệm giảng dạy và sự quen thuộc với hệ thống đánh giá dựa trên AI. Trước khi bước vào giai đoạn thu thập dữ liệu chính, chúng tôi đã tiến hành thử nghiệm mẫu để đảm bảo tính hợp lệ và độ tin cậy của phiếu khảo sát. Một nhóm nhỏ giáo viên IELTS được mời tham gia hoàn thành phiếu khảo sát để đưa ra phản hồi về tính rõ ràng, sự liên quan đến câu hỏi nghiên cứu và mức độ dễ hiểu của bảng khảo sát. Dựa trên các ý kiến quý báu này, chúng tôi đã thực hiện các sửa đổi cần thiết để cải thiện phiếu khảo sát và loại bỏ các lỗi gây ra sự khó hiểu. Sau đó, phiên bản cuối cùng của phiếu khảo sát đã được công bố công khai dưới dạng biểu mẫu Google để thu thập phản hồi. Sau khi đã thu thập 105 phản hồi, chúng tôi phân tích 21 phản hồi đáp ứng cả hai tiêu chí sau: 1) Những người đã là giáo viên dạy IELTS ít nhất 5 năm; 2) Những người đã sử dụng ChatGPT từ 6 tháng trở lên. Điều này sẽ đảm bảo nghiên cứu đưa ra kết quả đáng tin cậy.

**2.1.2. Phỏng vấn**

Quá trình thu thập dữ liệu định tính bao gồm việc tiến hành các cuộc phỏng vấn bán cấu trúc với một nhóm ba giáo viên IELTS (A, B, C) được lựa chọn cẩn thận: Giáo viên A (31 tuổi) đã có kinh nghiệm giảng dạy tiếng Anh hơn 9 năm tại một trung tâm Anh ngữ tư nhân; Giáo viên B (46 tuổi) có 15 năm kinh nghiệm dạy tiếng Anh tại các trung tâm Anh ngữ; Giáo viên C (53 tuổi) có 31 năm kinh nghiệm dạy tiếng Anh tại một trường công lập. Họ đã sử dụng ChatGPT để đánh giá bài viết IELTS trong hơn 6 tháng và tự nguyện chia sẻ quan điểm của mình. Các cuộc phỏng vấn này đi sâu vào quan điểm của giáo viên về việc sử dụng ChatGPT trong việc đánh giá các bài viết IELTS. Các cuộc phỏng vấn bán cấu trúc vừa có định hướng vừa linh hoạt nhờ vào sự cân bằng giữa các câu hỏi đặt ra sẵn và cuộc trò chuyện mở tùy theo câu trả lời của người tham gia phỏng vấn. Để đảm bảo tính nghiêm ngặt và đáng tin cậy của dữ liệu phỏng vấn, một giao thức phỏng vấn được phát triển tỉ mỉ. Giao thức này cung cấp một khung câu hỏi và gợi ý định sẵn nhằm định hướng cuộc phỏng vấn trong khi vẫn cho phép nhà nghiên cứu đào sâu những chủ đề mới nổi lên và khai thác sâu hơn các phản ứng của người tham gia. Với sự đồng ý của người tham gia, các cuộc phỏng vấn được ghi âm và ghi chép chính xác từng lời để được phân tích toàn diện. Các vấn đề đạo đức được xem trọng trong toàn bộ quá trình nghiên cứu. Người tham gia nhận được giải thích chi tiết về mục tiêu, quy trình và quyền lợi của người tham gia nghiên cứu trước khi đồng ý tham gia phỏng vấn. Hơn nữa, họ được đảm bảo rằng, sự tham gia của họ là hoàn toàn tự nguyện và họ có thể dừng lại bất cứ lúc nào mà không phải chịu hậu quả. Để bảo vệ tính bảo mật của người tham gia, các bí danh đã được sử dụng trong quá trình phân tích dữ liệu và báo cáo. Ngoài ra, tất cả dữ liệu thu thập được lưu trữ một cách an toàn và chỉ có thể truy cập bởi các nhà nghiên cứu.

**2.2. Kết quả nghiên cứu**

**2.2.1. Câu hỏi 1: Những thách thức chính mà các giáo viên IELTS gặp phải khi tích hợp ChatGPT vào việc đánh giá viết IELTS?**

*a. Kết quả khảo sát định lượng*

Thông qua phân tích chi tiết của phản hồi khảo sát định lượng và dữ liệu phỏng vấn định tính, chúng tôi nhận thấy được một loạt các thách thức. Dữ liệu định lượng từ khảo sát (xem Bảng 1) chỉ ra rằng, ChatGPT tập trung quá nhiều vào những lỗi ngữ pháp nhỏ trong khi bỏ qua những vấn đề quan trọng khác mà giáo viên cần ChatGPT xem xét sâu sắc hơn (trung bình = 3.52; độ lệch chuẩn = 0.814). Ngoài ra, ChatGPT phân tích, góp ý khía cạnh Task Achievement (trung bình = 3.95; độ lệch chuẩn = 0.973) thiếu chiều sâu. Những thách thức này do người dùng chưa tạo lập được câu lệnh phù hợp với công cụ ChatGPT. Hơn nữa, người tham gia khảo sát đều đồng tình rằng, ChatGPT gặp khó khăn trong việc cung cấp từ vựng phù hợp và tự nhiên trong các ngữ cảnh khác nhau do những hạn chế vốn có của các mô hình AI trong việc nắm bắt sự phức tạp của ngôn ngữ và bối cảnh văn hóa của con người (trung bình = 3.71; độ lệch chuẩn = 0.902). Cuối cùng, người tham gia khảo sát đồng ý với quan điểm rằng, ChatGPT đánh giá không nhất quán giữa các bài viết IELTS có chủ đề và phong cách viết khác nhau (trung bình = 3.71; độ lệch chuẩn = 0.902). Việc duy trì tính nhất quán trong các đánh giá là một thách thức nổi bật. Bởi vì tiêu chí đánh giá của ChatGPT dễ bị ảnh hưởng bởi cách sử dụng ngôn ngữ khác nhau, ngữ pháp và phong cách viết đa dạng, việc đảm bảo sự đồng nhất trong các tiêu chuẩn đánh giá của ChatGPT là một rào cản đáng kể.

*b. Kết quả phỏng vấn định tính*

“Tôi thường xuyên gặp khó khăn trong việc tinh chỉnh các câu lệnh để khai thác được phản hồi như mong đợi” (Giáo viên A). “ChatGPT thường không hiểu hoặc không cung cấp được từ ngữ, cấu trúc với sắc thái nghĩa đa dạng phù hợp với những hoàn cảnh và

**Bảng 1: Những thách thức mà giáo viên tham gia khảo sát gặp phải khi sử dụng ChatGPT**

Mệnh đề	1	2	3	4	5	Trung bình	Độ lệch chuẩn
ChatGPT quá tập trung vào các lỗi ngữ pháp nhỏ trong khi bỏ qua những lỗi lớn hơn.	3	2	7	7	2	3.52	0.814
ChatGPT gặp khó khăn trong việc cung cấp câu trả lời về cách sử dụng từ vựng linh hoạt tùy ngữ cảnh.	2	3	8	6	2	3.71	0.902
ChatGPT chưa thể phân tích sâu về khía cạnh Task Achievement của bài viết.	0	2	6	9	4	3.95	0.973
ChatGPT gặp khó khăn trong việc đánh giá một cách nhất quán các bài viết thuộc nhiều chủ đề và phong cách viết khác nhau.	0	2	4	8	7	3.71	0.902
Khó đánh giá chính xác điểm mạch lạc và liên kết (Coherence and Cohesion) khi dùng ChatGPT.	0	2	8	9	2	3.14	1.108

(Ghi chú: 1. Hoàn toàn không đồng ý với mệnh đề đã cho; 2. Không đồng ý với mệnh đề đã cho; 3. Trung lập; 4. Đồng ý với mệnh đề đã cho; 5. Hoàn toàn đồng ý với mệnh đề đã cho.)

văn hoá khác nhau” (Giáo viên B). “ChatGPT thường không hiểu chính xác ý của tôi nên đưa ra các câu trả lời không liên quan” (Giáo viên C).

Một thách thức đáng chú ý được các giáo viên tham gia phỏng vấn nhấn mạnh là việc tạo ra những câu lệnh cho ChatGPT một cách hiệu quả để công cụ AI này đưa ra các câu trả lời phù hợp với mong đợi của người dùng. Tất cả giáo viên tham gia phỏng vấn (A, B, và C) đều nhất trí bày tỏ khó khăn trong việc điều chỉnh các câu lệnh để thu được các phản hồi chi tiết và chính xác từ ChatGPT. Cụ thể, A và C đã nêu rõ hạn chế trong các phản hồi của ChatGPT. Họ nhấn mạnh việc cần phải điều chỉnh các câu lệnh nhiều lần mới có thể thu được các phản hồi như mong đợi từ ChatGPT. Ngoài ra, các giáo viên thể hiện lo ngại về việc ChatGPT hiểu lầm hoặc đơn giản hóa những khái niệm đa sắc thái. Điều này ảnh hưởng đến độ chính xác của việc đánh giá. Giáo viên B chỉ ra hạn chế của ChatGPT trong việc thấu hiểu văn hóa và ngữ cảnh thực tế trong đời sống con người.

**2.2.2. Câu hỏi 2: Giáo viên IELTS đã làm thế nào để đối phó và vượt qua những thách thức này?**

*a. Kết quả khảo sát định lượng*

Ở phần này, những câu hỏi khảo sát và phỏng vấn chủ trọng tìm hiểu các chiến lược được các giáo viên áp dụng để vượt qua những thách thức liên quan đến việc tích hợp ChatGPT vào đánh giá bài viết IELTS. Người tham gia khảo sát đề xuất sử dụng các câu hỏi “Tại sao” trong câu lệnh để khuyến khích ChatGPT phân

tích bài viết sâu hơn (trung bình = 3.71; độ lệch chuẩn = 0.717) (xem Bảng 2). Hơn nữa, dữ liệu định lượng khuyến khích chiến lược hợp tác với các giáo viên khác (trung bình = 3.52; độ lệch chuẩn = 0.814) và việc giáo viên tìm kiếm hỗ trợ kĩ thuật cũng như cập nhật phương pháp đánh giá tiên tiến (trung bình = 3.38; độ lệch chuẩn = 0.805). Dữ liệu khảo sát đề xuất thêm một số chiến lược khác. Cụ thể, các giáo viên đề xuất việc sử dụng bảng tiêu chí chấm điểm để đánh giá bài viết IELTS chính thức do ban tổ chức cuộc thi IELTS cung cấp để định hướng việc chấm điểm và đánh giá thay vì phụ thuộc vào ChatGPT một cách vô định (trung bình = 3.71; độ lệch chuẩn = 0.902). Mặt khác, giáo viên khuyến khích việc điều chỉnh thủ công các câu trả lời của ChatGPT để vừa tiết kiệm thời gian vừa tăng chất lượng đánh giá (trung bình = 3.95; độ lệch chuẩn = 0.921).

*b. Kết quả phỏng vấn định tính*

Những phát hiện định lượng trên được củng cố bởi chia sẻ của các giáo viên trong cuộc phỏng vấn.

“Theo kinh nghiệm của tôi, khi đặt câu hỏi “Tại sao”, ChatGPT sẽ đưa ra câu trả lời chi tiết hơn. Ví dụ, khi tôi đặt câu lệnh ‘Vui lòng chấm điểm bài viết IELTS này và lí giải đánh giá của bạn cho mỗi tiêu chí?’, ChatGPT đưa ra số điểm với giải thích ngắn gọn về điểm mạnh và điểm cần cải thiện cho mỗi tiêu chí trong tổng cộng 4 tiêu chí đánh giá. Tuy nhiên, khi tôi tiếp tục hỏi ‘Tại sao bạn lại cho điểm Task Achievement là 7.0?’, ChatGPT đưa ra câu trả lời chi tiết hơn về chiều sâu lập luận của học sinh, chỉ ra việc học sinh đã

**Bảng 2: Các chiến lược giáo viên sử dụng để vượt qua những thách thức**

Mệnh đề	1	2	3	4	5	Trung bình	Độ lệch chuẩn
Việc đặt câu hỏi “Tại sao” trong câu lệnh khuyến khích ChatGPT phân tích bài viết sâu hơn.	0	1	6	12	2	3.71	0.717
Hợp tác với các giáo viên dùng ChatGPT khác giúp ích cho việc trao đổi kinh nghiệm.	0	3	5	12	1	3.52	0.814
Giáo viên cần tìm kiếm hỗ trợ kĩ thuật và những cập nhật mới nhất.	0	3	8	9	1	3.38	0.805
Bảng tiêu chí chấm điểm và các ví dụ giúp làm rõ cách thức chấm điểm.	0	2	6	9	4	3.71	0.902
Giáo viên cần tự điều chỉnh thủ công câu trả lời từ ChatGPT để tiết kiệm thời gian và tăng hiệu quả của đánh giá.	1	0	3	12	5	3.95	0.921
Giáo viên cần kiểm tra câu trả lời từ ChatGPT trước khi chia sẻ với học sinh nhằm đảm bảo chất lượng.	1	0	6	8	6	3.86	1.014
Giáo viên cần hướng dẫn học sinh các từ nối phù hợp trong các ngữ cảnh đa dạng để giúp cải thiện tính mạch lạc của bài viết.	0	0	8	13	0	3.62	0.498
Tiếp xúc với các mẫu bài viết đa dạng mở rộng sự hiểu biết về từ vựng của học sinh.	0	1	6	12	2	3.71	0.717
Học sinh luyện tập những cấu trúc ngữ pháp được ưa chuộng trong cuộc thi IELTS giúp tăng điểm ngữ pháp.	1	3	6	9	2	3.38	1.024
Giáo viên so sánh các đánh giá giúp nhận diện và sửa chữa các khác biệt.	0	1	9	9	2	3.57	0.746

(Ghi chú: 1. Hoàn toàn không đồng ý với mệnh đề đã cho; 2. Không đồng ý với mệnh đề đã cho; 3. Trung lập; 4. Đồng ý với mệnh đề đã cho; 5. Hoàn toàn đồng ý với mệnh đề đã cho).

đưa ra luận điểm đối lập (counterargument) hay chưa, đánh giá cách học sinh kết luận. Sau đó, ChatGPT cũng đưa ra một kết luận mẫu có điểm Task Achievement cao hơn để học sinh tham khảo” (Giáo viên B).

“Tôi thấy những góp ý của thầy cô trong cộng đồng giáo viên online và tại trung tâm giúp tôi biết cách tạo lập và điều chỉnh câu lệnh hiệu quả” (Giáo viên A).

Các giáo viên tán thành với ý kiến sử dụng câu hỏi “Tại sao” trong câu lệnh giúp khai thác được những phân tích sâu hơn của ChatGPT. Giáo viên B đưa ra ví dụ cụ thể cho luận điểm này như trên (xem Hình 1 và Hình 2). Tiếp theo, những giáo viên đề cao việc tìm kiếm sự hỗ trợ từ cộng đồng giáo dục hoặc các chuyên gia AI để giải quyết các thách thức và tận dụng kiến thức chung của cộng đồng. Việc hợp tác với các giáo viên khác cũng sử dụng ChatGPT để đánh giá năng lực ngôn ngữ của học sinh giúp giáo viên chia sẻ các phương pháp tốt nhất, khắc phục sự cố để tạo môi trường thúc đẩy sự đổi mới và phát triển. Bên cạnh đó, việc liên tục phát triển chuyên môn là một trong các chiến lược quan trọng của giáo viên. Họ cần cập nhật những tiến bộ trong công nghệ AI và các phương pháp đánh giá tiên tiến.

I'll evaluate the essay based on the IELTS Task 2 criteria: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy.

**Task Response**

Score: 7

- **Strengths:** The essay clearly addresses the task prompt, presenting a well-rounded argument for why increased accessibility to international travel is positive. Examples are provided to support the main points.
- **Areas for Improvement:** The conclusion could be more developed, possibly summarizing the main points discussed in the essay. Additionally, addressing potential counterarguments could strengthen the response.

Hình 1: Ví dụ câu trả lời của ChatGPT trước khi đặt câu lệnh “Tại sao”

**2.2.3. Câu hỏi 3: Các giáo viên IELTS đề xuất gì để áp dụng ChatGPT vào việc đánh giá bài viết IELTS một cách hiệu quả?**

*a. Kết quả khảo sát định lượng*

Dữ liệu khảo sát thể hiện sự đồng thuận của các giáo viên về tầm quan trọng của việc xây dựng cộng đồng hợp tác (trung bình = 3.81; độ lệch chuẩn = 0.75), thiết lập các nguyên tắc đạo đức rõ ràng và thử nghiệm, nghiên cứu kỹ trước khi triển khai trên quy mô lớn để tạo điều kiện cho việc áp dụng ChatGPT để đánh giá bài viết IELTS một cách có trách nhiệm và hiệu quả (trung bình = 4.1; độ lệch chuẩn = 0.768). Việc đầu tư vào nghiên cứu và phát triển khoa học công nghệ AI cần được chú trọng (trung bình = 3.76; độ lệch chuẩn = 0.995).

I assigned a score of 7 for Task Response based on the following considerations:

**Strengths**

1. **Clear Position:** The essay clearly states a position that the trend of global integration has positive impacts.
2. **Relevant Examples:** Examples, such as the increased travel from Vietnam to Korea and outsourcing by companies like Apple, effectively support the main points.
3. **Coverage of Points:** The essay covers both individual and national/global impacts, demonstrating a comprehensive approach to the task.

**Areas for Improvement**

1. **Depth of Argumentation:** While the points are relevant and supported by examples, the essay could benefit from deeper analysis. For instance, more detailed explanations of how these developments impact individuals and economies would strengthen the argument.
2. **Counterarguments:** Addressing potential counterarguments or drawbacks of increased international travel could demonstrate a more balanced and critical understanding of the topic. For example, discussing potential negative impacts on local cultures or environments could add depth.
3. **Conclusion Development:** The conclusion restates the positive impacts but could be more developed. Summarizing the key points discussed in the body and reinforcing the overall argument would provide a stronger ending.

**Example of a More Developed Conclusion:**

“In conclusion, the increased accessibility and affordability of international travel have significant positive impacts on individuals by broadening their cultural perspectives and enhancing their career prospects. Economically, it facilitates global trade and creates job opportunities, contributing to the growth of both developed and developing countries. While there may be some challenges, such as cultural homogenization and environmental concerns, the overall benefits of this trend are substantial.”

These improvements would help achieve a higher score in the Task Response criterion.

Hình 2: Ví dụ câu trả lời của ChatGPT sau khi đặt câu lệnh “Tại sao”

**Bảng 3: Các đề xuất nhằm ứng dụng ChatGPT một cách hiệu quả**

Mệnh đề	1	2	3	4	5	Trung bình	Độ lệch chuẩn
Việc tạo ra cộng đồng hợp tác giữa các giáo viên để chia sẻ các phương pháp tốt nhất và giải quyết vấn đề là hữu ích.	0	1	5	12	3	3.81	0.750
Thiết lập nguyên tắc đạo đức rõ ràng cho việc sử dụng các công cụ AI trong đánh giá bài viết là cần thiết.	0	0	5	9	7	4.10	0.768
Thử nghiệm và đánh giá kĩ lưỡng trước khi sử dụng các công cụ AI trong các bài đánh giá quan trọng như kì thi IELTS.	0	0	4	12	5	4.05	0.669
Việc phát triển tài liệu đào tạo tiêu chuẩn về việc đánh giá IELTS bằng công cụ AI là rất quan trọng.	2	0	8	8	3	3.48	1.078
Đầu tư vào nghiên cứu và phát triển nhằm cải thiện khả năng đánh giá chính xác của ChatGPT.	0	2	7	6	6	3.76	0.995

(Ghi chú: 1. Hoàn toàn không đồng ý với mệnh đề đã cho; 2. Không đồng ý với mệnh đề đã cho; 3. Trung lập; 4. Đồng ý với mệnh đề đã cho; 5. Hoàn toàn đồng ý với mệnh đề đã cho)

*b. Kết quả phỏng vấn định tính*

“Tôi đề xuất phỏng học thuật của các trung tâm Anh ngữ tổng hợp bộ câu lệnh mẫu để giáo viên tham khảo và sử dụng khi chấm bài viết IELTS” (Giáo viên B).

“Việc hướng dẫn giáo viên cách sử dụng AI để chấm điểm bài viết IELTS là rất cần thiết. Một số ý tưởng tôi nghĩ đến ngay là nhà trường có thể tổ chức những buổi chia sẻ và những buổi tập huấn chuyên sâu để các thầy cô trao đổi kinh nghiệm và tài liệu sử dụng AI” và “Các cơ sở giáo dục cần nghiên cứu độ tin cậy và hiệu quả của việc sử dụng AI để đánh giá bài viết của học sinh trước khi phổ biến rộng rãi cho toàn bộ đội ngũ giáo viên” (Giáo viên C).

“Chúng ta nên xây dựng bộ nguyên tắc đạo đức khi sử dụng AI để tránh đạo văn và tránh việc giáo viên phụ thuộc hoàn toàn vào công cụ này. Tôi cho rằng, một trong những nguyên tắc hàng đầu là giáo viên cần điều chỉnh câu trả lời của ChatGPT dựa trên bảng tiêu chí chấm điểm tiêu chuẩn của IELTS và kinh nghiệm của bản thân thay vì sao chép hoàn toàn” (Giáo viên A).

Kết quả các cuộc phỏng vấn định tính ủng hộ những phát hiện định lượng. Giáo viên tham gia phỏng vấn nhấn mạnh sự cần thiết của việc xây dựng tài liệu đào tạo tiêu chuẩn để hướng dẫn giáo viên sử dụng ChatGPT hiệu quả. Các cơ sở giáo dục cần trang bị cho giáo viên đầy đủ kiến thức về việc đánh giá năng lực viết bằng AI nhờ vào cung cấp tài liệu như bộ câu lệnh tham khảo và các buổi đào tạo, buổi hội thảo thực hành.

Hơn nữa, các giáo viên được phỏng vấn nhấn mạnh giá trị của việc thúc đẩy cộng đồng hợp tác trong lĩnh vực giáo dục để tạo điều kiện cho việc chia sẻ kiến thức, hỗ trợ đồng nghiệp và giải quyết thắc mắc chung của cộng đồng. Bằng cách tạo ra mạng lưới của các giáo viên, nhà nghiên cứu và chuyên gia AI, các cơ sở giáo dục tận dụng kiến thức chung để giải quyết các thách thức phổ biến, chia sẻ các phương pháp tốt nhất và thúc đẩy sự đổi mới trong giảng dạy và đánh giá ngôn ngữ dựa trên AI.

Những yếu tố đạo đức cũng được quan tâm. Các giáo viên IELTS nhấn mạnh sự quan trọng của việc thiết lập những nguyên tắc đạo đức rõ ràng để định hướng việc sử dụng AI trong giáo dục bằng cách tuân thủ những nguyên tắc về sự minh bạch, trách nhiệm và công bằng, các giáo viên có thể đảm bảo rằng quy trình đánh giá dựa trên AI tuân thủ những tiêu chuẩn về tính chính trực và công bằng, bảo vệ lợi ích của học sinh và duy trì niềm tin của cộng đồng vào các cơ sở giáo dục.

Cuối cùng, giáo viên bày tỏ sự cần thiết của việc thử

nghiệm cẩn thận trước khi triển khai rộng rãi phương pháp đánh giá dựa trên AI. Bằng cách tiến hành các nghiên cứu đánh giá và xác thực nghiêm ngặt, các cơ sở giáo dục có thể đánh giá hiệu suất, độ tin cậy và tính hợp lệ của những công cụ đánh giá dựa trên AI, nhận diện nhiều hạn chế tiềm ẩn và hoàn thiện các chiến lược triển khai trước khi triển khai toàn diện cho đội ngũ giáo viên.

**3. Kết luận**

Kết quả nghiên cứu chỉ ra một số thách thức liên quan đến việc tích hợp ChatGPT, công cụ AI vào đánh giá kỹ năng viết IELTS. Các giáo viên đã nêu ra những vấn đề chính như khó khăn trong việc tạo lập câu lệnh với ChatGPT, hạn chế của ChatGPT trong việc hiểu ngữ cảnh văn hóa hoặc sự đơn giản hóa quá mức các khái niệm phức tạp ảnh hưởng đến độ chính xác của việc đánh giá và sự không đồng nhất trong đánh giá của ChatGPT đối với các bài viết có chủ đề và phong cách viết khác nhau. Ngoài ra, lo ngại về việc phụ thuộc quá nhiều vào ChatGPT và sự hoài nghi của học sinh đối với độ chính xác của việc đánh giá dựa trên AI đã được đề cập.

Để ứng phó với những thách thức trên, giáo viên đã sử dụng đa dạng các chiến lược khác nhau để tăng cường hiệu quả của ChatGPT trong việc đánh giá bài viết IELTS. Các chiến lược bao gồm điều chỉnh liên tục các câu lệnh với ChatGPT cho đến khi đạt kết quả như mong muốn, tìm kiếm sự hỗ trợ từ cộng đồng giáo dục và công nghệ, kiểm tra kỹ góp ý từ ChatGPT trước khi cho điểm cuối cùng và liên tục phát triển chuyên môn nghiệp vụ sư phạm và đánh giá. Những chiến lược này nhấn mạnh tầm quan trọng của việc hợp tác và phát triển chuyên môn trong bối cảnh công nghệ liên tục đổi mới như hiện nay.

Các giáo viên tham gia nghiên cứu đã đưa ra những khuyến nghị dành cho lãnh đạo các đơn vị giáo dục, các nhà hoạch định chính sách, nhà phát triển chương trình học, phát triển chương trình đào tạo giáo viên, và thiết kế đề thi để tối ưu hóa việc áp dụng ChatGPT trong việc đánh giá viết IELTS. Các khuyến nghị chính bao gồm việc phát triển các tài liệu và buổi đào tạo tiêu chuẩn, thúc đẩy cộng đồng hợp tác cho giáo viên, thiết lập các nguyên tắc đạo đức rõ ràng và thử nghiệm kỹ trước khi triển khai. Những khuyến nghị này cung cấp hướng dẫn thực tiễn cho các nhà giáo dục nhằm tăng cường hiệu quả và tính trung thực của các phương pháp đánh giá bằng công nghệ AI.

**Tài liệu tham khảo**

[1] Council, B. (n.d.), (2023), *What is IELTS?*, British Council, <https://takeielts.britishcouncil.org/take-ielts/what-ielts>.  
 [2] Văn Trang, (3/2024), *Cơ hội trúng tuyển đại học khi*

*sở hữu IELTS từ 4.5*, Báo Lao động, Hà Nội, <https://laodong.vn/giao-duc/co-hoi-trung-tuyen-dai-hoc-khi-so-huu-ielts-tu-45-1312230.ldo>.

- [3] Nguyen, H. V., Nguyen, M. X., & Dao, P. (2020), *The impact of IELTS on English language teachers in Central Vietnam*, IELTS Research Reports, Cambridge.
- [4] Quỳnh Khanh, (7/2023), *IELTS Writing band descriptors - Tiêu chí chấm IELTS Writing*, IDP IELTS, Việt Nam, <https://ielts.idp.com/vietnam/prepare/article-ielts-writing-band-descriptors>.
- [5] Chen, L., Chen, P., & Lin, Z. (2020), *Artificial Intelligence in Education: A Review*, IEEE Access, doi: 10.1109/ACCESS.2020.2988510.
- [6] Cleveland, D. (5/2023), *Here's how I used ChatGPT to assess student writing*, Ditch That Textbook. <https://ditchthattextbook.com/chatgpt-ai-assess-student-writing/>.
- [7] Academy, K. (2023), *Update: Introducing... Khanmigo! Khan Academy's AI Tool*, Khan Academy, USA. <https://support.khanacademy.org/hc/en-us/articles/14394953976333--Update-Introducing-Khanmigo-Khan-Academy-s-AI-Tool>.
- [8] Runyon, K. R. M., Montilus, K. D., Nachman, L., Herrick, K., & Ferrara, L. A. (2022), *ETS® AI Labs™ Ways of Working Tutorial: How to Build Evidence-Based, User-Obsessed, AI-Enabled Learning Solutions in an Agile Framework*, In *Lecture Notes in Computer Science*, pp.119-122, Springer Science+Business Media, [https://doi.org/10.1007/978-3-031-11647-6\\_21](https://doi.org/10.1007/978-3-031-11647-6_21).

## TEACHERS' VIEWS ON UTILIZING ARTIFICIAL INTELLIGENCE IN IELTS WRITING ASSESSMENT

Ho Nguyen Thai Oanh<sup>1</sup>, Van Pham Ky An\*<sup>2</sup>

<sup>1</sup> Email: oanhho.master@gmail.com  
 Van Lang University  
 45 Nguyen Khac Nhu street, District 1,  
 Ho Chi Minh City, Vietnam  
 \* Corresponding author  
<sup>2</sup> Email: anvan@alumni.upenn.edu  
 Pennsylvania University  
 3700 Walnut, Philadelphia,  
 Pennsylvania, America

**ABSTRACT:** *In the context of an increasing number of colleges utilizing the International English Language Testing System (IELTS) certification for admissions purposes, there is a growing demand for IELTS preparation. Consequently, teachers are compelled to allocate more time to assess IELTS writing tasks to meet the escalating needs of students. Amid the popularity of ChatGPT, an artificial intelligence (AI) model introduced in November 2022, many teachers have adopted ChatGPT for grading IELTS writing tasks to save time and enhance the assessment quality. However, not all educators effectively incorporate this new technology. This study delves into the perspectives of IELTS instructors regarding the challenges and optimal methodologies to utilize AI in language evaluation, particularly in assessing writing tasks. The study adopts both a quantitative survey and qualitative interviews to collect input from instructors. The findings indicate that the primary challenge is crafting comprehensive and meticulous prompts to enable ChatGPT to provide accurate assessments. The study recommends a method to create effective prompts and underscores the importance of building collaborative support networks among educators, providing in-depth training for educators in AI utilization, and establishing ethical guidelines for using AI in language assessment. The new findings will assist educators in optimizing the application of AI in English writing skill evaluation.*

**KEYWORDS:** Artificial Intelligence (AI), ChatGPT, IELTS, language assessment, EdTech.