

COMPARING CHATGPT'S FEEDBACK WITH TEACHER'S FEEDBACK IN REVISING FRENCH TEXTS

Do Thi Bích Thuý

Email: dbthuy2003@gmail.com

University of Languages and International Studies –
Vietnam National University, Hanoi
Pham Van Dong street, Cau Giay district,
Hanoi, Vietnam

Received: 25/02/2025

Revised: 11/4/2025

Accepted: 12/5/2025

Published: 15/6/2025

Abstract: This study compares ChatGPT's feedback with teacher feedback in revising B1+-level French writing for 47 university students. The data comprises questionnaires and students' reflective reports, focusing on the comparison between ChatGPT's and teacher's feedback in terms of the quantity and quality of feedback, motivational encouragement, grading criteria, clarity, and usefulness. The results indicate that the scores assigned by ChatGPT and teachers were relatively similar. Overall, teacher feedback was rated more favorably by students across most aspects; however, ChatGPT detected more errors and provided more suggestions for improving expression. A total of 51.06% of students preferred teacher feedback, while 48.94% favored a combination of both types. This study represents one of the pioneering efforts to explore the integration of ChatGPT into French language teaching in Vietnam.

Keywords: ChatGPT feedback, teacher feedback, writing instruction, student perceptions, comparison.

SO SÁNH PHẢN HỒI CỦA CHATGPT VỚI PHẢN HỒI CỦA GIÁO VIÊN TRONG SỬA BÀI VIẾT TIẾNG PHÁP

Đỗ Thị Bích Thủy

Email: dbthuy2003@gmail.com

Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội
Phạm Văn Đồng, quận Cầu Giấy,
Hà Nội, Việt Nam

Nhận bài: 25/02/2025

Chỉnh sửa xong: 11/4/2025

Chấp nhận đăng: 12/5/2025

Xuất bản: 15/6/2025

Tóm tắt: Nghiên cứu của chúng tôi so sánh phản hồi của ChatGPT với phản hồi của giáo viên trong sửa bài viết tiếng Pháp cấp độ B1+ cho 47 sinh viên đại học. Dữ liệu phân tích là bảng hỏi và các báo cáo chiêm nghiệm của sinh viên so sánh phản hồi của ChatGPT và của giáo viên về số lượng và chất lượng phản hồi, khả năng động viên khích lệ, các tiêu chí chấm bài viết, độ dễ hiểu, tính hữu dụng. Kết quả cho thấy, điểm chấm bài viết của ChatGPT và của giáo viên khá tương đương. Phản hồi của giáo viên được sinh viên đánh giá cao hơn phản hồi của ChatGPT ở hầu hết mọi phương diện, tuy nhiên ChatGPT nhận được nhiều lỗi hơn và đưa ra nhiều đề xuất chữa lỗi diễn đạt tốt. 51,06% sinh viên thích giáo viên nhận xét hơn, 48,94% thích kết hợp cả hai phương thức phản hồi. Nghiên cứu này là một trong những nghiên cứu đầu tiên về ứng dụng ChatGPT trong giảng dạy tiếng Pháp tại Việt Nam.

Từ khóa: Phản hồi của ChatGPT, phản hồi của giáo viên, dạy viết, đánh giá của sinh viên, so sánh.

1. Đặt vấn đề

Rất nhiều nghiên cứu đã khẳng định muốn viết tốt thì học sinh phải luyện tập viết và phải được sửa bài viết (Woodworth & Barkaoui, 2020). Tuy nhiên, đưa ra nhận xét chi tiết đầy đủ về mọi khía cạnh của một bài viết cũng như chấm điểm bài viết mất nhiều thời gian và công sức của giáo viên (Steiss và cộng sự, 2024). ChatGPT ra đời và được sử dụng như một giải pháp trợ giúp cho giáo viên trong việc sửa bài viết cho học sinh ở mọi cấp độ và ngôn ngữ. Miễn

phí, nhanh, linh hoạt, tương tác tốt, có tính cá thể hóa cao, khá chính xác, ChatGPT càng ngày càng được sử dụng rộng rãi trong dạy và học ngoại ngữ (Baskara, 2023). Câu hỏi đặt ra là, liệu ChatGPT có chiếm mất vị trí của giáo viên trong lớp học hay sinh viên vẫn cần tới hướng dẫn, phản hồi của giáo viên khi học viết?

Một số nghiên cứu gần đây đã so sánh hiệu quả phản hồi của ChatGPT và của giáo viên khi chữa bài viết ngoại ngữ cho học sinh. Fokides & Peristeraki

(2024) đã nghiên cứu trên bài viết của học sinh tiểu học, gồm 20 bài tiếng Anh tiếng nước ngoài và 20 bài tiếng Hi Lạp tiếng mẹ đẻ cấp độ A2-B1. Tình nguyện tham gia chấm bài cho nghiên cứu là 10 giáo viên tiểu học chấm bài tiếng Hi Lạp và 10 giáo viên người Hi Lạp dạy tiếng Anh chấm bài tiếng Anh. Câu lệnh yêu cầu ChatGPT đóng vai một giáo viên dạy tiếng Anh (hay một giáo viên dạy tiếng Hi Lạp cấp Tiểu học), sử dụng mọi kiến thức ChatGPT có về tiếng Anh (hay tiếng Hi Lạp) để chữa bài. ChatGPT sẽ phải lập một danh sách các lỗi ngữ pháp, từ vựng, chính tả; một danh sách nhận xét về điểm mạnh của bài viết (ngữ pháp, chính tả, diễn đạt, cấu trúc, nội dung); một danh sách các yếu điểm của bài viết cũng theo các tiêu chí kể trên. Sau đó, ChatGPT cần viết một bản tóm tắt các phản hồi để gửi cho người học, dựa trên danh sách lỗi đã tìm được, điểm mạnh và điểm yếu của bài viết. Dựa vào những phản hồi này, AI được yêu cầu chấm điểm bài viết trên thang từ 1-10. Kết quả cho thấy, ở bài tiếng Anh, phản hồi của ChatGPT tốt hơn phản hồi của giáo viên ở cả số lượng và chất lượng. Trí tuệ nhân tạo nhận được nhiều lỗi hơn, đưa ra các nhận xét chi tiết hơn, khen ngợi học sinh nhiều hơn. Tuy nhiên, ChatGPT có phong cách viết trịnh trọng và hàn lâm hơn giáo viên. Ở bài viết tiếng Hi Lạp, ChatGPT nhận xét không tốt bằng giáo viên. Một số từ và mệnh đề viết đúng nhưng ChatGPT báo là viết sai. Máy cũng không tập trung nhận xét lỗi về ngôn ngữ nhiều như giáo viên tiếng Hy Lạp. Theo Fokides & Peristeraki (2024), có vẻ như là Hi Lạp là một ngôn ngữ ít phổ thông hơn nên sẽ không được sử dụng nhiều để đào tạo AI, từ đó ảnh hưởng tới năng lực chữa bài của ChatGPT với tiếng Hi Lạp. Các nhà nghiên cứu cũng nhận thấy, ChatGPT mắc nhiều lỗi chính tả và ngữ pháp khi viết phản hồi về các bài viết tiếng Hy Lạp.

Steiss và cộng sự (2024) đã tiến hành một nghiên cứu so sánh chất lượng phản hồi của ChatGPT 3.5 và của giáo viên khi chữa bài luận môn lịch sử tại bang California, Mỹ. Chất lượng phản hồi được đo trên 5 yếu tố: 1) Nhận xét đưa ra dựa trên các tiêu chí chấm bài viết; 2) Có đưa ra định hướng giúp học sinh cải thiện bài viết; 3) Chính xác; 4) Ưu tiên các yếu tố quan trọng và có thể cải thiện được của bài viết; 5) Giọng điệu khích lệ động lực của học sinh. Tham gia nghiên cứu này có 200 học sinh và 16 người chấm bài đã có kinh nghiệm dạy viết lâu năm. Giáo viên được tập huấn về cách đưa ra phản hồi giúp người học cải thiện được năng lực, được cung cấp thang chấm và cách thức viết phản hồi cho các tiêu chí chấm cụ thể. Mỗi giáo viên dành khoảng 20-25 phút để chấm một bài và được trả thù lao khi tham gia chấm. Câu lệnh

yêu cầu ChatGPT đóng vai một giáo viên cấp hai, dùng cách nói thân thiện mang tính động viên, đưa ra 2-3 phản hồi cụ thể, nói rõ người học có thể cải thiện bài như thế nào (tương tự yêu cầu khi tập huấn giáo viên) và cho ví dụ nếu có thể. Kết quả cho thấy, giáo viên chấm tốt hơn ChatGPT ở mọi yếu tố, ngoại trừ yếu tố số 1 (nhận xét đưa ra dựa trên các tiêu chí chấm bài viết). Tuy nhiên, Steiss và cộng sự (2024, tr.7) khẳng định AI đưa ra phản hồi với chất lượng khá tốt. Khi chia bài viết thành 3 nhóm có kết quả cao, trung bình và thấp thì yếu tố độ chính xác và ưu tiên các yếu tố quan trọng có thể cải thiện được của bài viết của ChatGPT là thấp nhất với tệp bài kết quả cao và yếu tố giọng điệu khích lệ học sinh là thấp nhất với tệp bài kết quả thấp. ChatGPT không phát hiện ra khi người học nhầm nhân vật lịch sử (mặc dù nghiên cứu viên đã cung cấp các tài liệu cần thiết để viết bài cho AI khi đặt lệnh) hay trí tuệ nhân tạo tự mâu thuẫn với các phản hồi của chính mình (Steiss và cộng sự, 2024, tr.10).

Zeevy-Solovey (2024) đã so sánh phản hồi của bạn học, của ChatGPT và của giáo viên trong quá trình sửa bài viết tiếng Anh tiếng nước ngoài tại Israel. 30 sinh viên đại học được chia thành 15 cặp, viết bài luận, dùng bảng câu hỏi để sửa bài cho bạn mình (checklist) theo 5 tiêu chí: tổ chức bài viết, nội dung, ngữ pháp, hình thức văn bản (chính tả, dấu câu) và diễn đạt. Sau đó, các em viết lại bài lần 1 theo nhận xét của bạn rồi dùng ChatGPT sửa bài tiếp. Bảng câu hỏi với các tiêu chí để sửa bài cũng được gửi trong câu lệnh cho ChatGPT. Tiếp theo, sinh viên viết lại bài lần 2 theo phản hồi của ChatGPT. Cuối cùng, sinh viên nộp bài cho giáo viên chữa, cũng dùng checklist để nhận xét và đưa ra các gợi ý giúp người viết cải thiện chất lượng bài viết. Sau khi nhận được phản hồi từ giáo viên, sinh viên viết lại bài lần 3. Hết chu trình viết, sinh viên sẽ trả lời điều tra do giáo viên tiến hành nhằm so sánh hiệu quả của ba nguồn phản hồi: bạn học, ChatGPT và giáo viên. Kết quả phân tích trả lời điều tra cho thấy, 20% sinh viên nhận được phản hồi của ChatGPT ở tất cả 5 tiêu chí đánh giá bài viết; 20% sinh viên nhận phản hồi của AI về 3 tiêu chí là nội dung, diễn đạt và ngữ pháp; 20% sinh viên nhận phản hồi của AI về 2 tiêu chí hình thức văn bản và ngữ pháp. 93,33% người viết hiểu phản hồi của AI; 100% sinh viên đồng ý với phản hồi của AI và 100% thấy những phản hồi này là hữu ích. Chỉ có 6,67% sinh viên nhận được phản hồi của giáo viên về cả 5 tiêu chí sửa bài viết; 20% sinh viên nhận được phản hồi của giáo viên về 3 tiêu chí là nội dung, hình thức văn bản và ngữ pháp; 26,67% nhận phản hồi của giáo viên về hai tiêu chí

là hình thức văn bản và ngữ pháp. 100% sinh viên hiểu, đồng ý và thấy phản hồi của giáo viên là hữu ích. 46,67% sinh viên thích phản hồi của giáo viên vì nguồn nhận xét này được cho là đáng tin cậy nhất. 33,33% thích kết hợp cả phản hồi của ChatGPT và giáo viên vì AI có thể sửa bài được ngay và nhanh. Chỉ có một cặp đôi (6,67%) sinh viên thích ChatGPT hơn vì phản hồi nhanh, hữu dụng, nhận xét và đề xuất sửa bài rõ nhất, AI cũng biết động viên khích lệ người viết.

Tổng quan một số nghiên cứu trước đây cho thấy, đại đa số các nghiên cứu so sánh phản hồi của ChatGPT và phản hồi của giáo viên được tiến hành trên dữ liệu tiếng Anh là ngôn ngữ phổ biến nhất và thường được dùng để đào tạo AI nhất. Cũng giống như Fokides & Peristeraki (2024) muốn đo năng lực sửa bài của ChatGPT trên tiếng Hi Lạp, chúng tôi cũng muốn so sánh phản hồi của ChatGPT với phản hồi của giáo viên khi chữa bài tiếng Pháp, một thứ tiếng ít phổ biến hơn tiếng Anh. Nghiên cứu của chúng tôi có mục tiêu là so sánh phản hồi của ChatGPT và giáo viên khi sửa bài viết tiếng Pháp cấp độ B1 cho sinh viên năm thứ hai tiếng Pháp thông qua đánh giá định lượng kết hợp với định tính của người học. Câu hỏi nghiên cứu chúng tôi đặt ra là: Theo đánh giá của sinh viên, phản hồi từ ChatGPT có tốt như phản hồi của giáo viên khi sửa bài viết tiếng Pháp cấp độ B1+ hay không?

2. Phương pháp nghiên cứu

2.1. Mô tả thử nghiệm

Chúng tôi đã tiến hành thử nghiệm ứng dụng ChatGPT vào sửa bài viết luận cho 47 sinh viên ở hai lớp năm thứ hai, trình độ tiếng Pháp B1+ tại khoa Ngôn ngữ và Văn hóa Pháp, Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội. Các em bắt đầu học tiếng Pháp từ năm thứ nhất, như vậy đến thời điểm thử nghiệm, các em mới học tiếng Pháp được ba học kì. Trước thử nghiệm này thì sinh viên đã được làm quen với việc sử dụng ChatGPT trong quá trình học viết với những câu lệnh giúp tạo bài viết mẫu, phân tích các cấu trúc của bài viết mẫu, tạo bài tập luyện từ vựng, ngữ pháp.

Sau khi giáo viên chữa phần dàn ý và hướng dẫn cấu trúc bài luận, sinh viên viết bài trên giấy, không sử dụng công nghệ trợ giúp trong quá trình viết bài. Giáo viên chữa bài viết tay của sinh viên theo các tiêu chí chấm bài viết quy định trong đề cương môn học và trả lại bài cho sinh viên một tuần sau đó. Trong buổi trả bài, giáo viên đã tổng kết những lỗi sai nhiều sinh viên mắc phải ở cả hai lớp như chưa viết tốt phần mở bài và kết luận hay chưa biết

đưa các lập luận một cách rõ ràng. Giáo viên yêu cầu một số sinh viên viết tốt chia sẻ lên bảng phần mở bài và kết luận của mình cũng như các cách giới thiệu các ưu điểm và nhược điểm của một vấn đề. Sau đó, giáo viên đã bổ sung thêm bằng những ngữ liệu của mình. Tiếp theo, giáo viên đã làm demo trên máy chiếu cách đưa câu lệnh (bằng tiếng Pháp) cho ChatGPT chữa bài như sau:

##HƯỚNG DẪN: (Bạn là giáo viên tiếng Pháp tiếng nước ngoài giàu kinh nghiệm. Hãy chữa bài viết trong tệp đính kèm).

##NỘI DUNG: 1. Gạch chân tất cả các lỗi trong bài viết được đính kèm trong mục **##TÀI LIỆU**, giải thích bản chất của từng lỗi và đề xuất cách sửa. 2. Tóm tắt những điểm mạnh và điểm yếu của bài viết, đồng thời đưa ra lời động viên người viết. 3. Chấm điểm bài viết dựa trên các tiêu chí đánh giá. 4. Viết lại bài viết với trình độ ngôn ngữ B1+. 5. Đề xuất 20 câu trắc nghiệm từ vựng và 20 câu trắc nghiệm ngữ pháp nhằm khắc phục những điểm yếu về ngôn ngữ trong bài viết).

##YÊU CẦU: [Hãy chữa và chấm điểm bài viết theo yêu cầu của đề bài viết dành cho trình độ B1+ (Đề bài viết) và dựa theo các tiêu chí đánh giá (xem tài liệu Word đính kèm)).

##TÀI LIỆU: (bài viết, tiêu chí đánh giá).

Giáo viên cũng cung cấp cho sinh viên các tiêu chí chấm bài viết gồm hai phần:

Tiêu chí đánh giá gồm Mức độ hoàn thành bài (30%), Tổ chức bài viết (20%), Từ vựng (20%), Ngữ pháp (30%).

Thang điểm đánh giá cho từng tiêu chí với 4 mức Rất tốt, Tốt, Đạt yêu cầu, Không hoàn thành.

Sinh viên sử dụng phiên bản miễn phí của ChatGPT để chữa bài. Sinh viên được khuyến dùng bản đánh máy bài viết để gửi ChatGPT vì chúng tôi thấy rằng, ChatGPT nhận được nhiều lỗi hơn trên văn bản đánh máy so với văn bản dạng hình ảnh. Tuy nhiên, sinh viên cần lưu ý gõ lại đúng như bản viết tay vì nếu sinh viên tự sửa lỗi trong quá trình đánh máy thì kết quả chữa bài ở hai dạng văn bản sẽ khác nhau.

2.2. Thu thập và xử lý dữ liệu

Chúng tôi có gửi cho sinh viên một bảng khảo sát trên Google form để thu thập thông tin nhân khẩu học cũng như đánh giá định lượng của sinh viên về phản hồi của ChatGPT và của giáo viên. Bảng hỏi gồm có 13 câu hỏi, trong đó có 4 câu hỏi về giới tính, tuổi, năng lực viết, mức độ thành thạo công nghệ và

9 câu hỏi so sánh phản hồi của ChatGPT với phản hồi của giáo viên.

- 5. ChatGPT chấm cho em mấy điểm?
- 6. Giáo viên chấm cho em mấy điểm?
- 7. Ai nhận được nhiều lỗi hơn?
- 8. Ai nhận lỗi chính xác hơn?
- 9. Ai động viên khen ngợi em nhiều hơn?
- 10. Ai nhận xét dựa trên các tiêu chí chấm toàn diện hơn?
- 11. Ai nhận xét dễ hiểu hơn?
- 12. Nhận xét của ai hữu ích hơn?
- 13. Em thích được ai chữa bài hơn?

Tổng cộng 47 sinh viên ở cả hai lớp đã trả lời bảng hỏi. Chúng tôi đã dùng phần mềm thống kê định lượng SPSS phiên bản 26 để xử lý dữ liệu.

Sinh viên cũng được yêu cầu viết báo cáo chiêm nghiệm so sánh phản hồi của ChatGPT và của giáo viên theo các khía cạnh đã được đề cập từ câu hỏi số 7 đến 13 của bảng hỏi và đưa ra lí giải hay ví dụ minh họa cụ thể. Tổng cộng chúng tôi thu được 45 báo cáo chiêm nghiệm ở hai lớp. Mỗi báo cáo chiêm nghiệm dài khoảng 200 đến 500 từ được viết bằng tiếng Việt là ngôn ngữ mà sinh viên diễn đạt thuần thực hơn. Như vậy, chúng tôi sử dụng dữ liệu mà Le Boterf (2000) gọi là *verbalisation écrite différée* nghĩa là viết lại sau thực nghiệm những đánh giá và trải nghiệm của mình. Chúng tôi sử dụng phương pháp phân tích nội dung để nhóm các dữ liệu thu được về 7 chủ điểm chính, tương ứng với 7 gợi ý so sánh mà chúng tôi đã gửi cho sinh viên.

3. Kết quả nghiên cứu

Như chúng ta có thể thấy trong Bảng 1, đa số sinh viên tham gia thử nghiệm là nữ (39 sinh viên chiếm 83%), chỉ có 7 sinh viên nam và 1 sinh viên không muốn nêu cụ thể giới tính. Gần 90% sinh viên ở độ tuổi 19-20, chỉ có 6,3% là ở độ tuổi 21-23 và cá biệt có 4,2% (2 sinh viên) là ở độ tuổi 30-34. Trong câu hỏi về năng lực viết, chúng tôi có đưa ra 4 lựa chọn: Giỏi, Khá, Trung bình và Yếu nhưng không một sinh viên nào tự đánh giá mình ở mức Giỏi. Gần 20% sinh viên cho rằng, mình viết ở mức Khá, 68,1% các em nhận mình viết ở mức Trung bình và 12,8% số sinh viên được hỏi tự xếp mình có năng lực viết Yếu. Đại đa số sinh viên tự nhận xét có mức độ thành thạo công nghệ Khá, 10,6% ở mức Tốt và 27,7% ở mức Trung Bình.

So sánh điểm do máy và người chấm, chúng ta thấy điểm thấp nhất của máy là 3, của giáo viên là 4 điểm; điểm cao nhất của máy là 8,5, của giáo viên là 9 điểm. Điểm trung bình chung của máy chấm là 6,3 thấp hơn 0,3 so với người chấm là 6,6. Chúng tôi thống kê được 33 bài viết có điểm chênh từ 0 -1, chiếm 70% số lượng bài viết.

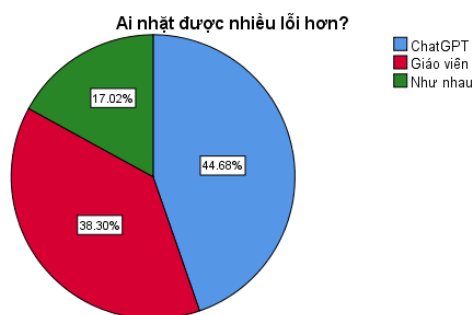
Hình 1 và Hình 2 cho thấy, ChatGPT vượt giáo viên về số lượng lỗi nhận được nhưng thua giáo viên về chất lượng phản hồi lỗi. Theo đánh giá của gần một nửa số sinh viên (44,68%) thì ChatGPT nhận được nhiều lỗi hơn giáo viên nhưng 53,19% thấy giáo viên nhận lỗi chính xác hơn. 38,3% sinh viên thấy giáo viên nhận được nhiều lỗi hơn trí tuệ nhân tạo trong khi chỉ có 2,13% cho rằng ChatGPT nhận lỗi đúng hơn.

Bảng 1: Thông tin nhân khẩu học của sinh viên tham gia thử nghiệm (số lượng 47)

Giới tính	Nam	7 (14,9%)	Nữ	39 (83%)	Không muốn nêu cụ thể	1 (2,1%)
Tuổi	19-20 tuổi	42 (89,4%)	21-23 tuổi	3 (6,3%)	30-34 tuổi	2 (4,2%)
Năng lực viết	Khá	9 (19,1%)	Trung bình	32 (68,1%)	Yếu	6 (12,8%)
Mức độ thành thạo công nghệ	Tốt	5 (10,6%)	Khá	29 (61,7%)	Trung bình	13 (27,7%)

Bảng 2: So sánh điểm chấm bài của ChatGPT và giáo viên

	Số lượng	Điểm thấp nhất	Điểm cao nhất	Trung bình	Độ lệch chuẩn
Điểm ChatGPT chấm	47	3.00	8.50	6.3	1.25732
Điểm giáo viên chấm	47	4.00	9.00	6.6	1.28858



Hình 1: So sánh số lượng nhật lỗi

“ChatGPT giải thích chi tiết và sửa lỗi đa dạng hơn” (E36).

“ChatGPT nhật nhiều lỗi hơn, cả lỗi nhỏ như dấu câu, từ vựng chưa tự nhiên. Giáo viên thì sửa trực tiếp vào bài, tập trung vào những lỗi quan trọng hơn” (E04).

Một số sinh viên đánh giá cao việc ChatGPT đưa ra các gợi ý sửa cách diễn đạt cho tốt hơn, đặc biệt là các sinh viên có năng lực viết khá.

“ChatGPT ko chỉ sửa lỗi sai mà còn sửa về ý tưởng, cách diễn đạt để câu văn trau chuốt hơn, hay hơn. Ví dụ cùng một ý như vậy nhưng ChatGPT sẽ gợi ý cách diễn đạt khác cho bài viết trở nên hay hơn” (E15).

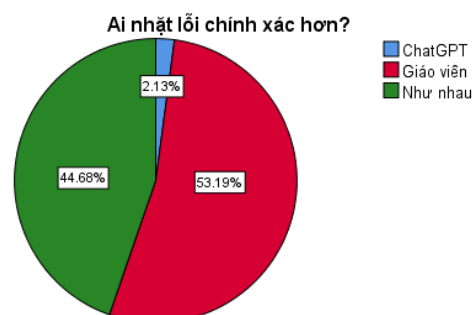
Trong báo cáo chiêm nghiệm, tất cả sinh viên được hỏi đều đánh giá ChatGPT là một công cụ sửa bài viết có độ chính xác cao. ChatGPT nhật rất tốt các lỗi chính tả, ngữ pháp, dấu câu. Tuy nhiên, hơn một nửa số sinh viên cho rằng, công cụ này còn mắc lỗi phản hồi sai do không hiểu được ngữ cảnh hay ý đồ của người viết như giáo viên.

“Nhưng khi xét đến lỗi diễn đạt hay cách triển khai ý tưởng, ChatGPT vẫn còn nhiều hạn chế. Nó chỉ có thể làm việc dựa trên dữ liệu lập trình sẵn mà không thực sự hiểu ngữ cảnh sâu sắc như giáo viên. Đối với giáo viên, bằng những kinh nghiệm và khả năng đọc hiểu tinh tế, có thể chỉ ra những lỗi sai mang tính logic, cách hành văn chưa tự nhiên hoặc ý tưởng chưa rõ ràng. Giáo viên không chỉ sửa lỗi mà còn hướng dẫn sinh viên cách tư duy, cách làm cho bài viết trở nên mạch lạc hơn” (E12).

Một số sinh viên cũng nêu rõ là ChatGPT đôi khi đưa ra những gợi ý chỉnh sửa quá đà (overcorrection).

“Nhiều trường hợp nhận xét cũng không quá cần thiết khi chủ ý của người dùng là như thế (mặc dù câu cú không sai cấu trúc ngữ pháp, không lỗi lặp từ...)” (E11).

Nhiều sinh viên khẳng định giáo viên chữa bài chính xác hơn, phát hiện lỗi tinh tế hơn, tôn trọng ý



Hình 2: So sánh chất lượng nhật lỗi

tưởng của người viết hơn hoặc đưa ra phân tích chữa lỗi mang tính cá nhân hóa hơn ChatGPT.

“Giáo viên với kinh nghiệm giảng dạy thực tế có khả năng phát hiện những lỗi tinh tế hơn, những lỗi liên quan đến văn phong. Vì vậy, về độ chính xác tổng thể, giáo viên vẫn nhỉnh hơn” (E21).

“AI sửa lỗi rất dài dòng nhiều khi khiến cho sinh viên cảm thấy chán và ko muốn đọc. Giáo viên nhật lỗi chính xác và giữ nguyên ý của bài viết” (E44).

“Giáo viên thường sẽ đưa ra thêm ví dụ để người học hiểu rõ hơn về cách dùng từ hoặc cấu trúc câu. Ví dụ, nếu học sinh sai về cách dùng mạo từ, giáo viên có thể cung cấp thêm vài câu ví dụ để minh họa sự khác biệt” (E09).

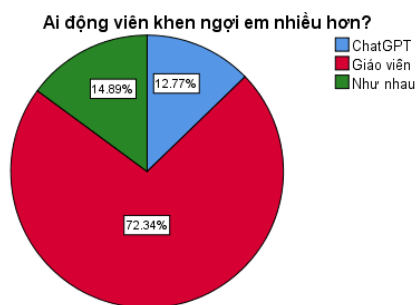
Gần 3/4 số sinh viên cho rằng, giáo viên động viên khích lệ người học nhiều hơn ChatGPT. 14,89% nhận xét AI và giáo viên động viên như nhau và chỉ có 12,77% sinh viên thấy ChatGPT động viên nhiều hơn giáo viên. Phản hồi của giáo viên thường mang tính cá nhân hóa cao hơn, có cảm xúc hơn, công nhận nỗ lực của người học nên thường được sinh viên thích hơn.

“Về mặt động viên, khen ngợi, những lời nhận xét của giáo viên mang lại cảm giác chân thực và khích lệ hơn. Em cảm nhận được rằng, thầy/cô đã đọc bài một cách nghiêm túc và ghi nhận nỗ lực của em” (E32).

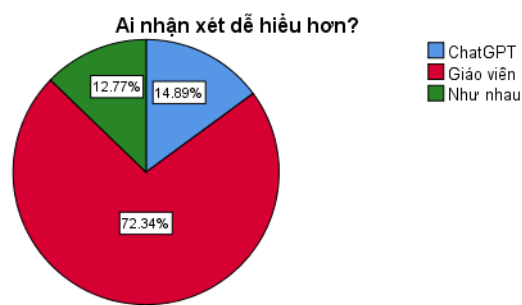
“ChatGPT có khích lệ nhưng không đáng kể. Cô rất hay khích lệ em và các bạn để chúng em tiến bộ hơn, chăm học hơn nên em rất thích” (E15).

“Chat GPT thường đưa ra nhận xét khá khách quan, trung lập. Tuy nhiên, những lời khen này có phần chung chung và thiếu tính cá nhân hóa. Ngược lại, giáo viên hiểu rõ sự tiến bộ của từng sinh viên nên có thể đưa ra những lời động viên phù hợp, cụ thể hơn, đánh giá thực tế và đôi khi đi kèm với sự công nhận về sự nỗ lực của sinh viên” (E21).

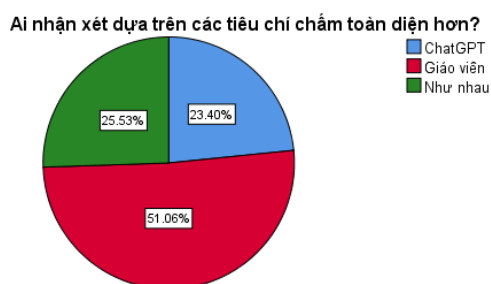
Có khoảng một nửa số sinh viên cho rằng, giáo viên nhận xét trên các tiêu chí chấm toàn diện hơn



Hình 3: So sánh về động viên, khen ngợi người viết



Hình 5: So sánh về độ dễ hiểu của nhận xét



Hình 4: So sánh về nhận xét trên các tiêu chí chấm



Hình 6: So sánh về tính hữu ích của nhận xét

trí tuệ nhân tạo. Tuy nhiên, gần 1/4 thấy ChatGPT nhận xét đầy đủ hơn giáo viên và gần 1/4 đánh giá hai nguồn phản hồi này là như nhau. Theo báo cáo chiêm nghiệm của sinh viên, ChatGPT sửa nhiều về lỗi ngôn ngữ nhưng hạn chế hơn giáo viên khi đưa ra phản hồi về ý tưởng, ví dụ minh họa, logic, cấu trúc bài viết, từ nối. Một nhược điểm của ChatGPT khi sửa lỗi ngôn ngữ là đôi khi không giải thích rõ lỗi sai hay đưa ra những từ và cấu trúc thay thế vượt quá năng lực ngôn ngữ của sinh viên.

“Khả năng đánh giá ý tưởng và ví dụ minh chứng của ChatGPT có thể bị hạn chế, đặc biệt là khi bài viết đòi hỏi sự sáng tạo và suy luận sâu sắc. Giáo viên có độ chính xác cao hơn trong việc nhận diện lỗi, đặc biệt là các lỗi liên quan đến ngữ cảnh và diễn đạt, có khả năng đánh giá sâu sắc hơn về ý tưởng, ví dụ minh chứng và tính logic của bài viết” (E22).

“Giáo viên không chỉ chấm bài dựa trên ngữ pháp và từ vựng, mà còn xem xét bài viết trong bối cảnh của đề bài, đánh giá tính hợp lý và logic. Ngược lại, ChatGPT chủ yếu đánh giá dựa trên việc so sánh với những bài viết chuẩn, tập trung nhiều vào việc sửa câu chữ để đáp ứng các quy tắc ngữ pháp hơn là tính mạch lạc toàn bài” (E30).

Ở tiêu chí dễ hiểu thì giáo viên vượt trội ChatGPT với 72,34% số sinh viên cho rằng, giáo viên nhận xét dễ hiểu hơn. 12,77% đánh giá người và máy nhận xét dễ hiểu tương đương và 14,89% thấy nhận xét của máy dễ hiểu hơn. Lí do của sự khác biệt lớn này là vì sau khi trả bài chữa thì giáo viên sẽ có một buổi tổng

kết những lỗi sai nhiều sinh viên mắc phải, giải thích và chữa cụ thể trên bảng cho cả lớp. Ngược lại, một số sinh viên thấy đôi khi câu trả lời của ChatGPT hơi dài và khó nắm bắt mặc dù các em có thể đối thoại với ChatGPT để yêu cầu giải thích rõ hơn.

“Em hiểu rõ nhận xét của cả hai nhưng em thấy giải thích của giáo viên dễ hiểu hơn vì có ví dụ cụ thể và chú thích về lỗi. Chat GPT tuy ngắn gọn nhưng đôi khi dùng thuật ngữ ngữ pháp khó hiểu”. (E05)

“ChatGPT diễn đạt khá rõ ràng, nhưng đôi khi nhận xét của nó hơi khô khan và thiếu sự giải thích sâu. Giáo viên có thể giải thích cặn kẽ hơn, đưa ra ví dụ minh họa cụ thể để người học dễ hiểu và áp dụng vào lần viết sau” (E03).

“Phản hồi của giáo viên thường ngắn gọn nhưng dễ hiểu. Nếu không hiểu, em có thể hỏi lại. ChatGPT có thể giải thích chi tiết nhưng đôi khi câu trả lời hơi dài và khó nắm bắt” (E45).

“ChatGPT thường sử dụng cấu trúc câu khá phức tạp, có thể ở trình độ B2. Điều này có thể khiến người học cảm thấy khó tiếp cận nếu chưa quen với kiểu văn phong này. Cô giáo có thể điều chỉnh mức độ phức tạp của câu để phù hợp với trình độ học viên hơn” (E36).

Về độ hữu ích, có hơn một nửa số sinh viên đánh giá phản hồi của giáo viên cao hơn, trong khi 38,3% đánh giá tương đương và chỉ có 6,38% đánh giá AI cao hơn. Sinh viên đánh giá hai nguồn phản hồi đều hữu ích nhưng theo cách khác nhau: Giáo viên giúp



Hình 7: So sánh về lựa chọn người chữa bài

hiểu sâu và cải thiện kỹ năng viết tổng thể (ý tưởng, cấu trúc, độ mạch lạc), còn Chat GPT hữu ích để phát hiện lỗi nhanh và thực hành sửa lỗi nhiều lần, đặc biệt là các lỗi về ngôn ngữ, diễn đạt.

“ChatGPT giúp sửa lỗi nhanh và chính xác nhưng giáo viên giúp phát triển tư duy viết và cải thiện nội dung bài viết toàn diện hơn” (E06).

“Cả hai loại nhận xét đều hữu ích theo cách riêng. ChatGPT giúp em cải thiện bài viết nhanh hơn, còn giáo viên giúp em hiểu sâu hơn về lỗi sai và cách sửa” (E13).

Tuy nhiên, có một sinh viên đánh giá ChatGPT hơn giáo viên ở độ hữu dụng của phản hồi. Em sinh viên này có tính kỷ luật cao và khả năng tự học khá tốt.

“ChatGPT hữu ích hơn vì vừa sửa lỗi, vừa giải thích, vừa đưa ra bài mẫu và bài tập luyện tập, giúp tiến bộ bền vững” (E36).

Hơn một nửa số sinh viên (51,06%) thích giáo viên chữa bài hơn và số lượng gần tương đương (48,94%) thích sử dụng cả hai nguồn phản hồi này để chữa bài. Rất nhiều sinh viên đánh giá cao khả năng sửa bài mọi lúc mọi nơi và rất nhanh chóng của ChatGPT, điều mà giáo viên không thể làm được. Tuy nhiên, điểm mạnh của giáo viên là cảm xúc, tính nhân văn, sự sâu sắc, tinh tế, tính cá thể hóa cao hơn.

“Cả hai cách chấm bài đều có điểm mạnh và điểm yếu. ChatGPT giúp phát hiện và sửa lỗi nhanh chóng nhưng giáo viên lại mang đến những đánh giá tinh tế và mang tính nhân văn hơn. Việc kết hợp cả hai phương pháp sẽ giúp học sinh cải thiện kỹ năng viết một cách toàn diện nhất” (E30).

“Em cũng nghĩ giữa việc cô chữa bài và ChatGPT chữa bài có sự khác biệt. Em thích được cô chữa bài hơn vì khi trao đổi trực tiếp với cô, em có thể hỏi lại những chỗ chưa rõ và nhận được những lời khuyên cụ thể về cách cải thiện. ChatGPT mặc dù rất nhanh và tiện, nhưng đôi khi em cảm thấy thiếu sự sâu sắc trong những lời giải thích, nhất là khi em cần một hướng dẫn chi tiết hơn về những lỗi mình gặp phải. Việc được cô chữa bài giúp em cảm thấy học

hỏi được nhiều hơn và tiến bộ nhanh hơn trong quá trình viết” (E35).

“Theo em, ChatGPT không thể thay thế giáo viên. Nhà giáo vẫn có một vai trò quan trọng trong việc giảng dạy và chữa bài. Giáo viên không chỉ chữa bài mà còn giúp học sinh hiểu và học hỏi từ những sai lầm của mình thông qua các cuộc trò chuyện, thảo luận và phản hồi trực tiếp. Giáo viên có thể chỉnh sửa bài linh hoạt hơn. Mỗi học sinh có khả năng học tập khác nhau, sự tiếp thu khác nhau mà ChatGPT không thể hiểu được. Giáo viên với vai trò là người dẫn dắt có thể dễ dàng hiểu được tình huống của từng học sinh” (E45).

4. Thảo luận

Chúng tôi đã thực nghiệm so sánh phản hồi chữa bài viết của giáo viên và ChatGPT với hai lớp sinh viên năm thứ hai tại một trường đại học chuyên ngoại ngữ tại Hà Nội. Kết quả cho thấy, điểm chấm bài viết giữa người và máy có sự khác biệt không quá lớn. Phân tích bảng hỏi và báo cáo chiêm nghiệm của các sinh viên sau thử nghiệm cho thấy, sinh viên đánh giá cao phản hồi của giáo viên hơn phản hồi của ChatGPT ở hầu hết mọi phương diện được hỏi, ngoại trừ số lượng phản hồi. Tuy nhiên, trong nghiên cứu của Fokides & Peristeraki (2024), khi chấm bài viết tiếng Anh, phản hồi của ChatGPT vượt trội phản hồi của giáo viên ở cả số lượng và chất lượng. Kết quả của chúng tôi đi cùng hướng với kết quả nghiên cứu của Fokides & Peristeraki (2024), của Steiss và cộng sự (2024) khi so sánh kết quả AI chấm bài và giáo viên chấm bài tiếng Hi Lạp và môn viết luận lịch sử cũng cho thấy giáo viên vượt trội AI ở hầu hết các tiêu chí. Có hai yếu tố có thể giải thích kết quả này. Thứ nhất, tiếng Pháp là một thứ tiếng ít phổ thông hơn tiếng Anh, ít được dùng để đào tạo AI hơn nên có thể năng lực ngôn ngữ của AI về tiếng Pháp không tốt bằng tiếng Anh. Thứ hai, giáo viên tham gia chấm tiếng Anh trong nghiên cứu của Fokides & Peristeraki (2024) không phải là giáo viên bản ngữ, trong khi giáo viên chấm tiếng Hi Lạp và chấm luận lịch sử là giáo viên bản ngữ. Trong nghiên cứu của chúng tôi, giáo viên không phải là người bản ngữ nhưng được đào tạo tiến sĩ về viết tại Pháp, có kinh nghiệm nghiên cứu và dạy viết lâu năm, tâm huyết với nghề.

Đánh giá của sinh viên cho biết ChatGPT dù nhặt lỗi về ngữ pháp, chính tả, từ vựng, dấu câu, hay sửa lỗi diễn đạt khá tốt nhưng vẫn mắc lỗi phản hồi sai hay lỗi chỉnh sửa quá đà do không hiểu ngữ cảnh hay ý muốn diễn đạt của người viết. Van Dis và cộng sự (2023) cũng đã chỉ ra rằng, Chatbot không

hiểu nghĩa mà chỉ học quy luật kết hợp theo ý nghĩa thống kê giữa các từ nên khi có những sự kết hợp ngoài quy luật thì máy sẽ báo lỗi. Steiss và cộng sự (2024), Su và cộng sự (2023), Zeevy-Solovey (2024) cho thấy, ChatGPT không nhất quán khi đưa ra phản hồi, hay tự mâu thuẫn với phản hồi của chính mình. Cũng giống như Yoon và cộng sự (2023), kết quả nghiên cứu của chúng tôi cho thấy ChatGPT hạn chế khi đưa ra phản hồi về tính mạch lạc, từ nối, độ tương thích của ý tưởng với đề bài.

Hơn một nửa số sinh viên thích giáo viên chữa bài cho mình hơn vì nhận xét của giáo viên có cảm xúc hơn, có tính cá nhân hóa cao hơn, dễ hiểu hơn, chính xác hơn, được giải thích lỗi rõ ràng hơn, được sửa nhiều về nội dung, ý tưởng, cấu trúc, minh chứng sát với yêu cầu của đề bài hơn. Người học trong nghiên cứu của Su và cộng sự (2023) cũng thích phản hồi của giáo viên hơn vì nó nằm trong vùng cận phát triển của học sinh hơn. Kết quả của chúng tôi tìm được cao hơn một chút so với 46,67% sinh viên trong nghiên cứu của Zeevy-Solovey (2024) thích nhận xét của giáo viên hơn nhận xét của ChatGPT và của bạn học. 48,94% số sinh viên mong muốn kết hợp cả phản hồi của ChatGPT và của giáo viên trong quá trình sửa bài: dùng ChatGPT để chữa ngay những lỗi ngôn ngữ, diễn đạt cơ bản sau đó nhờ giáo viên chữa để hoàn thiện hơn về ý tưởng, cấu trúc bài, văn phong, logic, tính hợp lý của ví dụ minh chứng. Tỷ lệ này cao hơn nghiên cứu của Zeevy-Solovey (2024), theo đó có 33,33% sinh viên thích kết hợp cả phản hồi của giáo viên và ChatGPT. Lý do là vì nghiên cứu của Zeevy-Solovey (2024) còn so sánh cả phản hồi của bạn học.

Tài liệu tham khảo

- Baskara, R. (2023). Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2), 343-358.
- Fokides, E., & Peristeraki, E. (2024). Comparing ChatGPT's correction and feedback comments with that of educators in the context of primary students' short essays written in English and Greek. *Education and Information Technologies*, 1-45.
- Le Boterf, G. (2000). *Construire les compétences individuelles et collectives*. Editions d'Organisation.
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., ... & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894.
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, 614(7947), 224-226. <https://doi.org/10.1038/d41586-023-00288-7>
- Woodworth, J. & Barkaoui, K. (2020). Perspectives on using automated writing evaluation systems to provide written corrective feedback in the ESL classroom. *TESL Canada Journal*, 37(2), 234-247.
- Yoon, S. Y., Miszoglad, E., & Pierce, L. R. (2023). *Evaluation of ChatGPT feedback on ELL writers' coherence and cohesion*. arXiv preprint arXiv:2310.06505.
- Zeevy-Solovey, O. (2024). Comparing peer, ChatGPT, and teacher corrective feedback in EFL writing: Students' perceptions and preferences. *Technology in Language Teaching & Learning*, 6(3), 1482-1482.

5. Kết luận

Cũng giống như kết quả của nhiều nghiên cứu trên các bài viết của học sinh, nghiên cứu của chúng tôi cho thấy sinh viên khi được hỏi khẳng định phản hồi của giáo viên ưu việt hơn phản hồi của ChatGPT về độ chính xác, có tính khích lệ người học cao hơn, dễ hiểu hơn, hữu ích hơn. Tuy nhiên, sinh viên, đặc biệt các em sinh viên khá cũng đánh giá cao ChatGPT với khả năng chấm và sửa bài viết tương đối chính xác gần như giáo viên, khả năng nhặt lỗi nhiều hơn giáo viên, khả năng đưa ra đề xuất sửa lỗi diễn đạt phong phú hơn giáo viên, tốc độ nhanh hơn giáo viên. Chúng tôi khuyến nghị nên ứng dụng ChatGPT vào dạy viết tiếng Pháp nói riêng và dạy ngôn ngữ nói chung để giảm tải áp lực chữa bài cho giáo viên, tận dụng tiềm năng đưa ra phản hồi có tính tương tác và có tính cá thể hóa của trí tuệ nhân tạo. Tuy nhiên, giáo viên vẫn đóng vai trò quyết định trong việc giúp người học lập dàn ý, tìm ý tưởng, tìm minh chứng phù hợp với yêu cầu của đề bài, đặt câu lệnh cho ChatGPT, kiểm soát chu trình viết sao cho sinh viên chỉ dùng AI để sửa bài chứ không dùng để viết bài. Cuối cùng, giáo viên cần chữa lại bài sau khi ChatGPT đã chữa để xử lý những lỗi mà máy sửa sai hay không tìm thấy. Việc tích hợp AI vào dạy học cần được thực hiện một cách có chọn lọc và có mục tiêu sư phạm rõ ràng, nhằm hỗ trợ chứ không thay thế quá trình học tập chủ động của người học.

Lời cảm ơn: Nghiên cứu này được cấp kinh phí bởi Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội trong khuôn khổ Nhiệm vụ KH&CN cấp Trường mã số N.25.07.