

ỨNG DỤNG CÔNG NGHỆ TRONG KIỂM TRA – ĐÁNH GIÁ NGOẠI NGỮ: TỪ TRẮC NGHIỆM TRUYỀN THỐNG ĐẾN AI CHẤM ĐIỂM TỰ ĐỘNG

Nguyễn Cẩm Nhung
Đại học Ngoại Ngữ, Đại học Quốc Gia Hà Nội

Tóm tắt: Trong bối cảnh chuyển đổi số giáo dục, kiểm tra – đánh giá ngoại ngữ đang có những thay đổi sâu sắc. Từ hình thức trắc nghiệm trên giấy truyền thống, các trường đại học tại Việt Nam đã chuyển dần sang thi trực tuyến và hiện nay đang bước vào giai đoạn ứng dụng trí tuệ nhân tạo (AI) trong chấm điểm tự động, đặc biệt ở kỹ năng viết và nói. Bài báo này phân tích tiến trình chuyển đổi từ phương pháp truyền thống sang công nghệ số, chỉ ra ưu điểm, hạn chế cũng như thách thức khi áp dụng AI. Đồng thời, bài báo đưa ra khuyến nghị nhằm xây dựng hệ thống đánh giá ngoại ngữ công bằng, hiệu quả và thích ứng với chuẩn mực quốc tế.

Từ khóa: kiểm tra – đánh giá, ngoại ngữ, công nghệ số, AI, chấm điểm tự động

APPLICATION OF TECHNOLOGY IN FOREIGN LANGUAGE TESTING AND ASSESSMENT: FROM TRADITIONAL MULTIPLE-CHOICE TO AI-BASED AUTOMATED SCORING

Nguyen Cam Nhung
University of Languages and International Studies, Vietnam National University, Hanoi

Abstract: In the context of digital transformation in education, foreign language assessment has undergone significant changes. From traditional paper-based tests, universities in Vietnam have gradually shifted to online examinations and are now entering a new phase with the application of artificial intelligence (AI) in automated scoring, particularly in writing and speaking skills. This article analyzes the transition from traditional to technology-enhanced assessment, highlighting its advantages, limitations, and challenges in the Vietnamese higher education context. It also provides recommendations for building a fair, effective, and internationally aligned language assessment system.

Keywords: assessment, foreign language, digital technology, AI, automated scoring

Nhận bài: 19/07/2025

Phản biện: 14/08/2025

Duyệt đăng: 19/08/2025

I. ĐẶT VẤN ĐỀ

Đánh giá kết quả học tập là một thành tố quan trọng trong giảng dạy ngoại ngữ, bởi nó không chỉ đo lường thành tích mà còn định hướng việc học (Brown & Abeywickrama, 2019). Trước đây, các kỳ thi ngoại ngữ tại Việt Nam chủ yếu dựa vào hình thức trắc nghiệm giấy hoặc thi trực tiếp trên lớp. Tuy nhiên, cùng với sự phát triển của công nghệ và yêu cầu đổi mới kiểm tra – đánh giá theo chuẩn quốc tế, phương pháp này đã có nhiều thay đổi.

Trong những năm gần đây, các trường đại học tại Việt Nam đã bắt đầu áp dụng thi trực tuyến thông qua các hệ thống quản lý học tập (LMS) như Moodle hoặc Google Classroom. Đồng thời, sự xuất hiện của trí tuệ nhân tạo mở ra triển vọng chấm điểm tự động, giúp giảm gánh nặng cho giảng viên và nâng cao tính khách quan trong đánh giá (Chapelle & Chung, 2010).

II. NỘI DUNG NGHIÊN CỨU

2.1. Khái niệm kiểm tra – đánh giá trong dạy và học ngoại ngữ

Kiểm tra – đánh giá (assessment) trong giảng dạy ngoại ngữ được hiểu là quá trình thu thập thông tin về năng lực ngôn ngữ của người học nhằm mục đích đo lường, phản hồi, và định hướng

học tập (Brown & Abeywickrama, 2019). Theo Weir (2005), một hệ thống đánh giá hiệu quả cần đảm bảo ba yếu tố: độ tin cậy (reliability), tính giá trị (validity) và tính công bằng (fairness). Trong bối cảnh hiện đại, kiểm tra – đánh giá không chỉ nhằm mục đích phân loại hay cấp chứng chỉ, mà còn đóng vai trò hỗ trợ người học tiến bộ thông qua phản hồi liên tục (Black & Wiliam, 2009).

Có ba loại hình đánh giá thường được nhắc đến: Đánh giá vì học tập (Assessment for Learning – AfL): nhấn mạnh vai trò phản hồi để cải thiện quá trình học.

Đánh giá của học tập (Assessment of Learning – AoL): thường là các kỳ thi tổng kết để xác nhận kết quả.

Đánh giá như học tập (Assessment as Learning – AaL): xem việc tự đánh giá là một phần trong quá trình học, giúp người học phát triển khả năng tự điều chỉnh (Earl, 2013).

Trong bối cảnh giáo dục đại học Việt Nam, các hình thức AfL và AaL đang ngày càng được quan tâm khi nhà trường hướng tới mô hình học tập lấy người học làm trung tâm và hội nhập quốc tế.

2.2. Sự chuyển đổi từ kiểm tra truyền thống sang công nghệ số

Trong nhiều thập kỷ, kiểm tra ngoại ngữ tại Việt Nam chủ yếu dựa vào hình thức trắc nghiệm giấy hoặc phỏng vấn trực tiếp. Phương pháp này có ưu điểm là quen thuộc, dễ triển khai, nhưng nhược điểm là tốn nhiều nhân lực, thiếu linh hoạt và khó quản lý dữ liệu (Nguyen & Pham, 2021).

Sự phát triển của Computer-Based Testing (CBT) đã tạo ra bước chuyển quan trọng. CBT cho phép tổ chức thi trên máy tính với ngân hàng câu hỏi phong phú, đồng thời giảm thiểu sai sót trong chấm điểm (Redecker & Johannessen, 2013). Kế tiếp, Computer-Adaptive Testing (CAT) ra đời với khả năng điều chỉnh mức độ khó của đề thi theo năng lực người học, tiêu biểu là kỳ thi GRE và GMAT. CAT giúp rút ngắn thời gian làm bài nhưng vẫn đảm bảo độ tin cậy (Wainer, 2000).

Trong giảng dạy ngoại ngữ, thi trực tuyến và CBT đã được áp dụng rộng rãi trong giai đoạn COVID-19. Tuy nhiên, sự xuất hiện của AI scoring systems mới thực sự mang tính cách mạng, bởi chúng có khả năng chấm các kỹ năng phức tạp như viết và nói, vốn trước đây phụ thuộc hoàn toàn vào giám khảo con người (Zhang, 2020).

2.3. Ứng dụng trí tuệ nhân tạo (AI) trong kiểm tra – đánh giá ngoại ngữ

AI trong đánh giá ngôn ngữ được ứng dụng theo hai hướng chính:

Xử lý ngôn ngữ tự nhiên (Natural Language Processing – NLP): cho phép hệ thống phân tích bài viết dựa trên ngữ pháp, từ vựng, mạch lạc và tính sáng tạo. Ví dụ: hệ thống ETS e-rater được sử dụng trong kỳ thi TOEFL iBT Writing (Ramineni & Williamson, 2018).

Nhận dạng và xử lý giọng nói (Speech Recognition): giúp chấm điểm kỹ năng nói dựa trên độ chính xác phát âm, ngữ điệu và mức độ lưu loát. Duolingo English Test hay Pearson PTE là minh chứng tiêu biểu (Flor & Maxwell, 2017).

Ưu điểm của AI là khả năng tự động hóa quy mô lớn, giảm thiên vị chủ quan từ giám khảo, và phản hồi tức thì cho người học. Tuy nhiên, một số nghiên cứu cảnh báo rằng AI vẫn có thể thiên lệch với những giọng nói, phương ngữ hoặc phong cách diễn đạt ngoài chuẩn (Zhang, 2020). Vì vậy, nhiều hệ thống hiện nay áp dụng hybrid scoring, kết hợp giữa AI và giám khảo con người để đảm bảo độ chính xác và công bằng.

2.4. Xu hướng quốc tế và tác động đối với Việt Nam

Trên thế giới, việc ứng dụng AI trong đánh giá ngoại ngữ đang trở thành xu hướng tất yếu, gắn liền với các kỳ thi chuẩn hóa quốc tế. Ở Việt Nam, việc hội nhập giáo dục và yêu cầu chuẩn ngoại ngữ trong khung năng lực 6 bậc (VSTEP) khiến việc áp dụng công nghệ trong kiểm tra ngày càng cấp thiết (Le & Tran, 2020).

Các nghiên cứu gần đây tại Việt Nam cho thấy sinh viên và giảng viên nhìn chung đánh giá tích cực về tính tiện lợi và minh bạch của thi trực tuyến. Tuy nhiên, họ cũng bày tỏ lo ngại về gian lận thi cử, độ tin cậy của hệ thống, và tính công bằng khi sử dụng AI (Nguyen & Pham, 2021). Điều này đặt ra nhu cầu cấp bách cho các trường đại học trong việc xây dựng hạ tầng công nghệ, đào tạo năng lực số và chuẩn hóa quy trình triển khai.

2.5. Thực tiễn ứng dụng tại Việt Nam

2.5.1. Giai đoạn trắc nghiệm truyền thống

Trong nhiều năm, hình thức thi trắc nghiệm trên giấy là phương thức chủ đạo trong kiểm tra – đánh giá ngoại ngữ tại Việt Nam, đặc biệt trong các kỳ thi quy mô lớn như TOEIC, VSTEP hoặc kỳ thi tốt nghiệp THPT quốc gia. Ưu điểm của phương thức này là dễ tổ chức, chi phí thấp và quen thuộc với cả giảng viên lẫn người học. Tuy nhiên, nó cũng bộc lộ những hạn chế rõ rệt: tốn thời gian chấm điểm, khó khai thác dữ liệu để phân tích kết quả học tập, và quan trọng hơn là thiếu khả năng đánh giá toàn diện các kỹ năng giao tiếp như nói và viết (Nguyen & Pham, 2021). Do đó, nhiều nhà nghiên cứu và quản lý giáo dục cho rằng việc duy trì hình thức truyền thống không còn đáp ứng được yêu cầu đổi mới trong bối cảnh hội nhập quốc tế.

2.5.2. Giai đoạn chuyển đổi sang kiểm tra trực tuyến

Bước ngoặt quan trọng xảy ra trong giai đoạn dịch COVID-19 (2020–2021) khi hầu hết các trường đại học buộc phải áp dụng hình thức thi trực tuyến. Các hệ thống quản lý học tập (LMS) như Moodle, Canvas, hoặc Google Classroom được tích hợp để tổ chức thi, kèm theo các công cụ giám sát như Zoom hoặc Google Meet. Ở một số trường, đề thi được thiết kế theo dạng ngân hàng câu hỏi trắc nghiệm ngẫu nhiên để hạn chế gian lận.

Việc chuyển sang thi trực tuyến mang lại nhiều lợi ích:

Tiện lợi và tiết kiệm thời gian: Sinh viên có thể làm bài ở bất cứ đâu, giảng viên không mất công chấm tay.

Khả năng phân tích dữ liệu: Hệ thống có thể thống kê kết quả, chỉ ra những câu hỏi khó/dễ để điều chỉnh đề thi.

Tính linh hoạt: Dễ dàng thay đổi hình thức đề thi, từ trắc nghiệm đến tự luận ngắn.

Tuy nhiên, các nghiên cứu cũng chỉ ra những thách thức lớn: gian lận học thuật, sự không đồng bộ về hạ tầng công nghệ giữa các trường, và tâm lý e ngại của cả giảng viên lẫn sinh viên trước những thay đổi đột ngột (Nguyễn & Phạm, 2021). Điều này cho thấy việc ứng dụng công nghệ trong đánh giá tại Việt Nam ban đầu mang tính “ứng phó” nhiều hơn là một chiến lược dài hạn.

2.5.3. Giai đoạn thử nghiệm trí tuệ nhân tạo trong chấm điểm

Trong bối cảnh giáo dục toàn cầu tiến tới tự động hóa chấm điểm bằng AI, một số cơ sở giáo dục đại học tại Việt Nam đã bắt đầu thử nghiệm áp dụng công nghệ này trong các kỹ năng viết và nói. Các phần mềm như ETS Criterion(chấm Writing) hoặc công cụ phân tích giọng nói dựa trên Speech Recognition được đưa vào sử dụng thử nghiệm. Một số trường kết hợp sử dụng các kỳ thi chuẩn hóa quốc tế (như Duolingo English Test hoặc Pearson PTE) để so sánh và tham khảo mô hình chấm điểm tự động.

Ứng dụng AI mang lại ba lợi ích nổi bật:

Tự động hóa quy mô lớn: Hàng nghìn bài viết hoặc bài nói có thể được chấm trong thời gian ngắn.

Giảm thiểu thiên vị chủ quan: Hệ thống đưa ra tiêu chí đồng nhất, hạn chế sự khác biệt giữa các giám khảo.

Phản hồi nhanh chóng: Sinh viên nhận kết quả ngay lập tức, có thể điều chỉnh chiến lược học tập.

Tuy vậy, thực tiễn triển khai cũng cho thấy nhiều rào cản:

Chưa đồng bộ về hạ tầng công nghệ: Nhiều trường đại học chưa có hệ thống máy chủ đủ mạnh hoặc mạng Internet ổn định để áp dụng rộng rãi.

Độ tin cậy chưa tuyệt đối: AI có thể đánh giá chưa chính xác đối với những giọng nói vùng miền, cách diễn đạt ngoài chuẩn hoặc lối viết sáng tạo.

Vấn đề bảo mật dữ liệu: Lưu trữ và xử lý dữ liệu giọng nói, bài viết của sinh viên cần tuân thủ chuẩn mực an toàn thông tin, vốn chưa được quy định rõ ràng tại Việt Nam.

Năng lực số của giảng viên và sinh viên: Một bộ phận chưa thành thạo trong sử dụng hệ thống, gây khó khăn cho việc triển khai.

2.5.4. Đánh giá chung

Nhìn tổng thể, quá trình ứng dụng công nghệ trong kiểm tra – đánh giá ngoại ngữ tại Việt Nam

đang ở giai đoạn chuyển tiếp: từ mô hình truyền thống sang thi trực tuyến và bắt đầu thử nghiệm AI. Đây là xu hướng tất yếu, phù hợp với định hướng chuyển đổi số quốc gia và yêu cầu hội nhập giáo dục. Tuy nhiên, để công nghệ thực sự phát huy hiệu quả, cần có một chiến lược dài hạn bao gồm: đầu tư hạ tầng, chuẩn hóa quy trình, nâng cao năng lực số cho giảng viên – sinh viên, và nghiên cứu nghiêm túc về tính công bằng, độ tin cậy của AI trong bối cảnh ngôn ngữ – văn hóa Việt Nam

2.6. Thách thức đặt ra

Mặc dù công nghệ và trí tuệ nhân tạo đã mở ra nhiều cơ hội đổi mới kiểm tra – đánh giá ngoại ngữ, thực tiễn tại Việt Nam vẫn còn tồn tại nhiều thách thức lớn cần được giải quyết. Có thể phân tích thành bốn nhóm chính sau:

2.6.1. Độ tin cậy và tính công bằng trong chấm điểm

Một trong những mối quan tâm hàng đầu là độ tin cậy (reliability) và tính công bằng (fairness) của hệ thống chấm điểm tự động. Các nghiên cứu quốc tế chỉ ra rằng AI có xu hướng ưu ái những dạng ngôn ngữ “chuẩn hóa” và gặp khó khăn khi xử lý giọng nói vùng miền hoặc phong cách viết sáng tạo (Flor & Maxwell, 2017; Zhang, 2020). Điều này đặt ra nguy cơ bất bình đẳng trong đánh giá, đặc biệt đối với người học đến từ các vùng có phương ngữ khác biệt.

2.6.2. Hạn chế về hạ tầng công nghệ

Việc triển khai thi trực tuyến và chấm điểm bằng AI đòi hỏi hệ thống máy chủ mạnh, băng thông Internet ổn định và phần mềm quản lý dữ liệu hiệu quả. Tuy nhiên, nhiều trường đại học Việt Nam vẫn gặp khó khăn về cơ sở vật chất, dẫn đến tình trạng nghẽn mạng, gián đoạn thi cử hoặc mất dữ liệu.

2.6.3. Năng lực số của giảng viên và sinh viên

Sự thành công của công nghệ đánh giá phụ thuộc nhiều vào năng lực số (digital literacy) của người sử dụng. Một bộ phận giảng viên chưa quen với việc thiết kế đề thi trên hệ thống LMS hoặc giám sát thi trực tuyến. Về phía sinh viên, tâm lý lo lắng khi làm bài trên nền tảng công nghệ mới cũng ảnh hưởng đến kết quả thi.

2.6.4. Vấn đề đạo đức và bảo mật dữ liệu

Ứng dụng AI trong đánh giá đồng nghĩa với việc lưu trữ khối lượng lớn dữ liệu giọng nói, bài viết, thậm chí là hình ảnh của người học. Hiện nay, các quy định pháp lý và chuẩn mực đạo đức về bảo mật dữ liệu tại Việt Nam chưa đầy đủ, dẫn đến nguy cơ rò rỉ hoặc sử dụng sai mục đích. Đây là vấn đề cấp thiết cần có cơ chế quản lý chặt chẽ.

2.7. Giải pháp và khuyến nghị

Để khắc phục những thách thức nêu trên, trong bài báo này đề xuất một số giải pháp chiến lược mang tính hệ thống:

2.7.1. Kết hợp đa dạng phương pháp đánh giá

Thay vì phụ thuộc hoàn toàn vào AI, các trường nên áp dụng mô hình hybrid scoring – kết hợp chấm điểm tự động với chấm điểm thủ công. Điều này vừa tận dụng được tốc độ và tính khách quan của AI, vừa bảo đảm được độ chính xác thông qua sự giám sát của giám khảo con người.

2.7.2. Nâng cao năng lực số cho giảng viên và sinh viên

Các chương trình tập huấn định kỳ về thiết kế đề thi trực tuyến, quản lý thi, và sử dụng công cụ AI là cần thiết. Song song đó, sinh viên cũng cần được hướng dẫn kỹ năng làm bài thi trên nền tảng công nghệ để giảm áp lực tâm lý và tăng sự tự tin.

2.7.3. Đầu tư hạ tầng và chuẩn hóa hệ thống

Nhà trường cần ưu tiên đầu tư hệ thống LMS hiện đại, có khả năng tích hợp trí tuệ nhân tạo, đảm bảo băng thông ổn định và tính bảo mật cao. Về phía cơ quan quản lý, Bộ Giáo dục và Đào tạo cần ban hành bộ chuẩn quốc gia về ứng dụng công nghệ và AI trong kiểm tra ngoại ngữ để thống nhất toàn hệ thống.

2.7.4. Tăng cường nghiên cứu và thử nghiệm

Các trường đại học nên triển khai các dự án thí điểm về chấm điểm tự động, tiến hành đánh giá so sánh giữa AI và giám khảo con người, từ đó điều chỉnh thuật toán cho phù hợp với đặc điểm ngôn ngữ – văn hóa của người học Việt Nam. Những

kết quả nghiên cứu này sẽ là cơ sở để mở rộng áp dụng trên quy mô lớn.

2.7.5. Bảo đảm an toàn và đạo đức trong sử dụng dữ liệu

Cần xây dựng quy định pháp lý rõ ràng về bảo mật dữ liệu, bao gồm lưu trữ, xử lý và chia sẻ. Các trường phải cam kết minh bạch trong việc sử dụng dữ liệu của người học, đồng thời áp dụng các biện pháp kỹ thuật để hạn chế rủi ro an ninh mạng.

III. KẾT LUẬN

Quá trình đổi mới kiểm tra – đánh giá ngoại ngữ tại Việt Nam đang trải qua những bước chuyển quan trọng: từ trắc nghiệm giấy truyền thống sang thi trực tuyến, và hiện nay là thử nghiệm AI chấm điểm tự động. Mỗi giai đoạn đều mang đến những lợi ích nhất định nhưng cũng đặt ra thách thức cần giải quyết.

Ứng dụng công nghệ, đặc biệt là trí tuệ nhân tạo, mở ra triển vọng nâng cao độ tin cậy, tính khách quan và hiệu quả trong đánh giá ngoại ngữ. Tuy nhiên, để công nghệ này thực sự phát huy hiệu quả tại Việt Nam, cần một chiến lược toàn diện bao gồm: kết hợp nhiều phương pháp đánh giá, nâng cao năng lực số, đầu tư hạ tầng, chuẩn hóa quy trình và bảo đảm đạo đức trong quản lý dữ liệu.

Trong tương lai, khi các giải pháp này được triển khai đồng bộ, hệ thống kiểm tra – đánh giá ngoại ngữ tại Việt Nam có thể tiệm cận chuẩn mực quốc tế, góp phần nâng cao chất lượng đào tạo ngoại ngữ và đáp ứng yêu cầu hội nhập toàn cầu.

TÀI LIỆU THAM KHẢO

- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices (3rd ed.)*. Pearson.
- Chapelle, C. A., & Chung, Y. R. (2010). The promise of NLP and speech processing technologies in language assessment. *Language Testing*, 27(3), 301–315.
- Flor, M., & Maxwell, M. (2017). Automatic scoring of speaking: State of the art, current practices, and future directions. *ETS Research Report Series*, 2017(1), 1–23.
- Ramineni, C., & Williamson, D. M. (2018). Understanding automated scoring: How do human and computer raters compare? *Assessing Writing*, 36, 1–14.
- Redecker, C., & Johannessen, Ø. (2013). Changing assessment—Towards a new assessment paradigm using ICT. *European Journal of Education*, 48(1), 79–96.
- Zhang, Y. (2020). Artificial intelligence in language testing and assessment: Opportunities and challenges. *Language Testing in Asia*, 10(1), 1–14.