

PHƯƠNG PHÁP KIỂM TRA THÍCH ỨNG (ADAPTIVE TESTING) TRONG ĐÁNH GIÁ KỸ NĂNG TIẾNG ANH CỦA SINH VIÊN

Trần Diệp Anh

Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội

Tóm tắt: Trong bối cảnh nhu cầu học và đánh giá năng lực tiếng Anh ngày càng cao tại các trường đại học, phương pháp kiểm tra thích ứng (Adaptive Testing) nổi lên như một giải pháp đổi mới giàu tiềm năng. Thay vì áp dụng cùng một bài kiểm tra cho tất cả người học, kiểm tra thích ứng cho phép hệ thống lựa chọn câu hỏi phù hợp với năng lực thực tế của từng cá nhân, từ đó tạo ra kết quả đánh giá sát thực và công bằng hơn. Với cơ sở lý thuyết vững chắc như Mô hình đáp ứng câu hỏi (IRT), phương pháp này không chỉ cải thiện độ chính xác mà còn góp phần nâng cao trải nghiệm người học nhờ tính cá nhân hóa và hiệu quả thời gian. Bài viết tập trung phân tích nguyên lý hoạt động, các dạng bài áp dụng trong giảng dạy tiếng Anh, lợi ích nổi bật đối với sinh viên, đồng thời đề xuất một số hướng triển khai phù hợp với điều kiện giáo dục tại Việt Nam.

Từ khóa: kiểm tra thích ứng, kỹ năng tiếng Anh, mô hình IRT

UNIVERSITY OF LANGUAGES AND INTERNATIONAL STUDIES, VIETNAM
NATIONAL UNIVERSITY, HA NOI

Tran Diep Anh

Email: trandiepanh2106@gmail.com

Abstract: In the context of the growing demand for English language learning and assessment at universities, adaptive testing has emerged as a promising and innovative solution. Instead of applying the same test to all learners, an adaptive test enables the system to select questions that match the actual proficiency level of each individual, thereby producing more accurate and equitable assessment outcomes. Based on a solid theoretical foundation such as the Item Response Theory (IRT), this method not only enhances measurement precision but also improves the learner's experience through personalization and time efficiency. This article analyzes the core principles of adaptive testing, explores its applicable formats in English language instruction, highlights its benefits for university students, and proposes several recommendations for effective implementation in the context of Vietnamese higher education.

Keywords: adaptive testing, English language skills, IRT model

Nhận bài: 12/05/2025

Phản biện: 16/06/2025

Duyệt đăng: 20/06/2025

I. ĐẶT VẤN ĐỀ

Trong bối cảnh đổi mới giáo dục đại học và chuyển đổi số mạnh mẽ tại Việt Nam, việc tích hợp công nghệ vào đánh giá ngôn ngữ không còn là xu hướng mà đã trở thành một yêu cầu tất yếu. Kiểm tra thích ứng vốn đã phát triển mạnh ở các quốc gia có nền giáo dục tiên tiến, đang dần được nghiên cứu và thử nghiệm trong môi trường đại học Việt Nam như một giải pháp đánh giá thông minh, cá nhân hóa và chính xác hơn.

Thay vì áp dụng cùng một bài kiểm tra cho toàn bộ người học, kiểm tra thích ứng dựa trên nguyên lý chọn câu hỏi phù hợp với năng lực thực tế của từng cá nhân, từ đó phản ánh đúng trình độ và tiềm năng của người học (Nguyen, 2024). Nghiên cứu của Bùi et al. (2024) về hệ thống kiểm tra từ vựng thích ứng CARVAT đã chỉ ra rằng mô hình này mang lại độ tin cậy cao và phù hợp với sinh viên không chuyên ngữ tại Việt Nam. Đồng thời, Duong et al. (2024) cũng cho thấy rằng đội ngũ giảng viên bày tỏ thái độ tích cực với phương pháp kiểm tra thích ứng, song vẫn còn tồn tại nhiều rào cản như hạ tầng kỹ thuật, nhận thức và chính sách hỗ trợ.

Trên cơ sở đó, bài viết này hướng đến việc làm rõ khái niệm và nguyên lý của kiểm tra thích ứng, phân tích các dạng bài kiểm tra phổ biến trong giảng dạy tiếng Anh, nêu bật lợi ích đối với sinh viên, đồng thời đề xuất một số khuyến nghị nhằm nâng cao hiệu quả ứng dụng phương pháp này trong môi trường giáo dục đại học tại Việt Nam.

II. NỘI DUNG NGHIÊN CỨU

2.1. Phương pháp nghiên cứu

Bài viết sử dụng phương pháp nghiên cứu tổng hợp tài liệu kết hợp phân tích định tính để làm rõ khái niệm, nguyên lý và ứng dụng của kiểm tra thích ứng trong giảng dạy tiếng Anh. Các nguồn tài liệu được lựa chọn gồm nghiên cứu trong và ngoài nước, đặc biệt là các công trình gần đây liên quan đến mô hình IRT và thực tiễn triển khai adaptive testing tại Việt Nam. Đồng thời, bài viết cũng phân tích một số trường hợp điển hình nhằm rút ra nhận định và đề xuất phù hợp với bối cảnh giáo dục đại học hiện nay.

2.2. Khái niệm và nguyên lý của kiểm tra thích ứng

Kiểm tra thích ứng (Adaptive Testing) hay chính xác hơn là kiểm tra thích ứng trên máy tính (Computerized Adaptive Testing - CAT) là một phương pháp đánh giá hiện đại trong đó câu hỏi được lựa chọn dựa trên năng lực trả lời của người làm bài ở từng thời điểm. Thay vì trình bày một bộ câu hỏi cố định như trong kiểm tra truyền thống, CAT sử dụng thuật toán để chọn ra câu hỏi kế tiếp phù hợp nhất với trình độ của thí sinh, nhằm tối ưu hóa độ chính xác trong việc ước lượng năng lực cá nhân (Wainer, 2000).

Nền tảng lý thuyết chủ đạo của kiểm tra thích ứng là Lý thuyết đáp ứng câu hỏi - Item Response Theory (IRT). Khác với Lý thuyết kiểm tra cổ điển (Classical Test Theory), IRT cho phép mô hình hóa mối quan hệ giữa xác suất trả lời đúng một câu hỏi và đặc điểm của câu hỏi đó (độ khó, độ phân biệt, xác suất đoán mò) cùng với năng lực ẩn (latent ability) của người học (Embretson & Reise, 2000). Nhờ đó, hệ thống có thể liên tục cập nhật ước lượng năng lực của thí sinh sau mỗi câu trả lời và điều chỉnh độ khó câu hỏi tiếp theo một cách linh hoạt.

Với nguyên lý này, kiểm tra thích ứng không chỉ rút ngắn thời gian làm bài mà còn tăng độ tin cậy và tính cá nhân hóa trong đánh giá. Mỗi thí sinh có một hành trình kiểm tra riêng biệt, không bị giới hạn bởi cấu trúc đề thi cứng nhắc, từ đó phản ánh chân thực hơn năng lực sử dụng ngôn ngữ của từng cá nhân (Weiss, 2011).

2.3. Các dạng bài kiểm tra thích ứng trong giảng dạy tiếng Anh

Trong giảng dạy tiếng Anh như một ngoại ngữ, kiểm tra thích ứng được ứng dụng linh hoạt vào nhiều kỹ năng ngôn ngữ khác nhau. Mỗi dạng bài đều khai thác ưu điểm của kiểm tra thích ứng là khả năng điều chỉnh độ khó câu hỏi phù hợp với năng lực cá nhân, từ đó giúp phản ánh chính xác trình độ thực tế của người học.

Một trong những dạng bài phổ biến nhất là kiểm tra từ vựng và ngữ pháp, thường áp dụng cho kiểm tra đầu vào (placement test). Ví dụ điển hình là hệ thống CARVAT phát triển tại Việt Nam, sử dụng công nghệ CAT để đánh giá vốn từ một cách chính xác và tiết kiệm thời gian (Bùi et al., 2024). Ngoài ra, các kỹ năng đọc hiểu và nghe cũng đặc biệt phù hợp với kiểm tra thích ứng nhờ tính phân tầng về độ khó của văn bản và nội dung nghe.

Một số hệ thống quốc tế đã tích hợp kiểm tra thích ứng vào toàn bộ quá trình đánh giá kỹ năng ngôn ngữ như Duolingo English Test (DET) - nơi mà toàn bộ bài thi (từ từ vựng, ngữ pháp đến nói, viết) đều được cá nhân hóa theo thời gian thực (Li, 2020). Bên cạnh đó, TOEFL iBT hiện nay cũng có một số phần kiểm tra điều chỉnh theo năng lực, đánh dấu bước chuyển dần sang mô hình kiểm tra linh hoạt hơn (ETS, 2023).

Nhiều quốc gia đã sớm triển khai kiểm tra thích ứng trong giảng dạy tiếng Anh như một phần của chính sách đổi mới đánh giá ngôn ngữ. Ở Hà Lan, kỳ thi Staatsexamen NT2 sử dụng mô hình thích ứng để đánh giá năng lực tiếng Hà Lan của người học ngoại ngữ, từ đó cung cấp mô hình tham chiếu cho việc phát triển các bài thi tiếng Anh tương tự. Tại Hàn Quốc, nhiều trường đại học như Korea University đã áp dụng kiểm tra thích ứng trong kiểm tra đầu vào môn tiếng Anh học thuật cho sinh viên năm nhất. Ở Mỹ, một số bang đã tích hợp CAT vào chương trình ESL công lập để đánh giá học sinh trung học không sử dụng tiếng Anh làm ngôn ngữ thứ nhất, nhằm xác định chính xác mức độ hỗ trợ cần thiết trong từng kỹ năng. Những mô hình này chứng minh rằng kiểm tra thích ứng không chỉ giới hạn trong các kỳ thi quốc tế, mà còn được các hệ thống giáo dục quốc gia sử dụng như một công cụ đánh giá chính thức trong bối cảnh đa dạng.

Sự đa dạng của các dạng bài kiểm tra thích ứng cho thấy phương pháp này không bị giới hạn trong bất kỳ kỹ năng nào, mà có thể tích hợp vào nhiều bối cảnh khác nhau như kiểm tra chuẩn đầu ra, đánh giá quá trình học, hoặc hỗ trợ tự học.

2.4. Lợi ích của phương pháp kiểm tra thích ứng trong đánh giá kỹ năng tiếng Anh của sinh viên

Thứ nhất, kiểm tra thích ứng (CAT) cung cấp độ chính xác đánh giá vượt trội nhờ khả năng tối ưu hóa lượng câu hỏi mỗi cá nhân cần làm. Lazendic và Martin (2018) chỉ ra rằng CAT tạo ra “lower achievement error rates,” tức sai số thấp hơn và đánh giá năng lực chính xác hơn so với phương pháp kiểm tra truyền thống. Điều này đặc biệt quan trọng trong bối cảnh đánh giá năng lực tiếng Anh, nơi sai số có thể dẫn đến phân loại sai trình độ sinh viên.

Thứ hai, CAT giúp giảm đáng kể thời gian thi, vì hệ thống chỉ đưa ra lượng câu hỏi vừa đủ để xác định năng lực. Mặc dù giảm số câu hỏi, hệ thống

vẫn đảm bảo độ tin cậy cao nhờ thuật toán chọn lọc thông minh (Lazendic & Martin, 2018). Kết quả là sinh viên ít mệt mỏi hơn, trải nghiệm thi tập trung và thoải mái hơn, đồng thời nhà trường cũng giảm tải tổ chức thi.

Thứ ba, hệ thống CAT thường tích hợp phản hồi tức thì cùng báo cáo chi tiết về kỹ năng, giúp cả người học và giảng viên theo dõi tiến bộ theo từng lĩnh vực: từ vựng, ngữ pháp, đọc, nghe. Adeoye et al. (2025) cũng nhấn mạnh rằng công cụ đánh giá thích ứng làm tăng rõ rệt sự gắn kết (engagement) và khả năng ghi nhớ kiến thức (knowledge retention) so với kiểm tra truyền thống tại Nigeria. Nhờ vậy, chương trình giảng dạy có thể được cá nhân hóa hơn.

Thứ tư, CAT được chứng minh là tăng động lực học tập và tự tin nhờ “self-efficacy” được củng cố khi người học liên tục trải qua các câu hỏi vừa sức. Lazendic và Martin (2018) ghi nhận rằng học sinh lớn tuổi (trong nghiên cứu của họ) báo cáo mức độ động lực cao hơn và sự gắn kết tốt hơn trong CAT so với kiểm tra truyền thống. Điều này rất quan trọng cho sinh viên đại học, bởi tăng động lực và tự tin chính là chìa khóa để thúc đẩy học tập chủ động.

Thứ năm, kiểm tra thích ứng phù hợp với xu thế chuyển đổi số trong giáo dục và mobile learning. Thiết kế trên nền tảng kỹ thuật số cho phép tổ chức kỳ thi linh hoạt, bất cứ lúc nào, bất cứ đâu qua thiết bị cá nhân. Adeoye et al. (2025) đã thí nghiệm ở bối cảnh đa trường tại Nigeria và thấy rõ rằng học sinh cảm thấy thoải mái, thuận tiện khi sử dụng adaptive assessment online so với kiểm tra truyền thống.

2.5. Một số khuyến nghị nâng cao hiệu quả ứng dụng kiểm tra thích ứng trong đánh giá kỹ năng tiếng Anh của sinh viên

Để kiểm tra thích ứng thực sự phát huy vai trò như một công cụ đánh giá ngôn ngữ hiện đại, chính xác và cá nhân hóa, các cơ sở giáo dục cần triển khai nhiều giải pháp đồng bộ về công nghệ, học thuật, nhân lực và chính sách.

Thứ nhất, cần xây dựng chiến lược tích hợp kiểm tra thích ứng vào hệ thống đánh giá học thuật một cách bài bản và có lộ trình. Việc áp dụng CAT không thể diễn ra đột ngột hay đơn lẻ. Các trường đại học cần xác định rõ mục tiêu của việc ứng dụng CAT (kiểm tra đầu vào, đánh giá giữa kỳ, chuẩn đầu ra...), lựa chọn kỹ năng phù hợp để

triển khai thử nghiệm (ví dụ: từ vựng hoặc đọc hiểu), từ đó mở rộng dần sang các kỹ năng còn lại. Quá trình triển khai cần có sự đồng hành của các đơn vị quản lý học vụ và bộ phận công nghệ thông tin để đảm bảo tính bền vững.

Thứ hai, việc phát triển ngân hàng câu hỏi theo chuẩn thích ứng là yếu tố then chốt quyết định hiệu quả của hệ thống. Một ngân hàng câu hỏi dùng cho kiểm tra thích ứng không chỉ cần phong phú về nội dung và hình thức, mà còn phải được gắn mã tham số (parameterized) theo các yếu tố như độ khó, độ phân biệt, và khả năng đoán mò theo mô hình IRT. Việc xây dựng một ngân hàng câu hỏi chuẩn đòi hỏi thời gian, chuyên môn cao và phải trải qua các bước hiệu chỉnh (pilot testing), phân tích thống kê, và cập nhật định kỳ. Đây là nhiệm vụ dài hạn nhưng bắt buộc nếu muốn kiểm tra thích ứng phản ánh đúng năng lực ngôn ngữ của người học.

Thứ ba, đào tạo đội ngũ giảng viên và cán bộ kỹ thuật là điều kiện tiên quyết. CAT không chỉ là một công cụ kỹ thuật mà còn là một phương pháp đánh giá mang tính sư phạm. Do đó, giảng viên cần được tập huấn để hiểu nguyên lý vận hành của hệ thống, cách xây dựng câu hỏi đúng chuẩn IRT, cũng như cách phân tích và sử dụng kết quả để phục vụ quá trình dạy - học. Song song đó, bộ phận kỹ thuật cần đảm bảo khả năng hỗ trợ triển khai CAT từ khâu vận hành hệ thống, xử lý lỗi, đến bảo mật dữ liệu thi.

Thứ tư, cần đầu tư vào hạ tầng công nghệ ổn định, thân thiện và tương thích. Kiểm tra thích ứng chỉ phát huy tối đa hiệu quả khi được triển khai trên nền tảng công nghệ đáng tin cậy: phần mềm đánh giá có khả năng chọn câu hỏi theo thời gian thực, hệ thống máy chủ đủ mạnh để xử lý đồng thời nhiều bài thi, giao diện rõ ràng, dễ sử dụng cho cả người học và giảng viên. Ngoài ra, việc kết nối hệ thống kiểm tra thích ứng với nền tảng quản lý học tập (LMS) như Moodle, Canvas hay hệ thống điểm số của nhà trường sẽ tạo ra một chu trình đánh giá - phản hồi - điều chỉnh hoàn chỉnh.

Thứ năm, kết quả kiểm tra thích ứng cần được tích hợp vào quy trình giảng dạy và tư vấn học tập. CAT không nên chỉ dừng lại ở việc “xếp loại” sinh viên, mà cần đóng vai trò như một công cụ học tập phản hồi, giúp người học hiểu rõ điểm mạnh, điểm yếu của mình, còn giảng viên thì có

cơ sở để cá nhân hóa nội dung giảng dạy. Ví dụ, sinh viên có điểm yếu ở phần từ vựng nên được đề xuất lộ trình học phù hợp, trong khi người có điểm mạnh ở phần đọc hiểu có thể được phân vào nhóm học nâng cao. Nếu kết quả CAT được sử dụng để điều chỉnh chương trình đào tạo và hỗ trợ học tập cá nhân, giá trị của nó sẽ vượt xa chức năng đánh giá thông thường.

III. KẾT LUẬN

Phương pháp kiểm tra thích ứng đang mở ra một hướng tiếp cận đầy hứa hẹn trong việc đánh giá kỹ năng tiếng Anh của sinh viên đại học. Không chỉ nâng cao độ chính xác và tính công bằng trong đánh giá, kiểm tra thích ứng còn góp phần tạo ra trải nghiệm thi cá nhân hóa, giảm áp lực thi cử và tăng cường hiệu quả học tập. Trên cơ sở lý thuyết vững chắc như IRT và nền tảng công nghệ phát triển mạnh, phương pháp này hoàn toàn có thể được tích hợp vào hệ thống đào tạo đại học tại Việt Nam, đặc biệt trong bối cảnh chuyển đổi số và nhu cầu nâng cao chuẩn đầu ra ngày càng cao. Tuy nhiên, việc triển khai kiểm

tra thích ứng trên thực tế hiện vẫn còn gặp một số hạn chế nhất định. Đó là thách thức trong việc xây dựng ngân hàng câu hỏi đạt chuẩn, thiếu nguồn lực kỹ thuật, và đặc biệt là nhận thức chưa đầy đủ từ phía giảng viên cũng như sinh viên về lợi ích của phương pháp này. Bên cạnh đó, phần lớn các nghiên cứu tại Việt Nam mới dừng lại ở giai đoạn thử nghiệm quy mô nhỏ, chưa có nhiều đánh giá hệ thống hoặc mô hình triển khai toàn diện trong các trường đại học.

Trong thời gian tới, cần có thêm các nghiên cứu thực nghiệm quy mô lớn hơn để đo lường tác động cụ thể của kiểm tra thích ứng đối với kết quả học tập và động lực học tiếng Anh của sinh viên. Đồng thời, các hướng nghiên cứu liên ngành, kết hợp giữa ngôn ngữ học, công nghệ giáo dục và khoa học dữ liệu sẽ giúp tối ưu hóa thuật toán chọn câu hỏi, cá nhân hóa trải nghiệm thi và nâng cao năng lực phản hồi tự động của hệ thống. Đây sẽ là nền tảng quan trọng để đưa kiểm tra thích ứng trở thành một phần tất yếu trong chiến lược đánh giá và đào tạo hiện đại ở bậc đại học.

TÀI LIỆU THAM KHẢO

- Adeoye, M. A., Prastikawati, E. F., & Riwayatningsih, R. (2025). Beyond traditional testing: Exploring the efficacy of adaptive assessments in shaping future learners. *Journal of Nonformal Education*, 11(1), 153-168.
- Bùi, T. K. P., Thanh, N. Q., Hùng, L. T., & Hà, N. T. (2024). Computerized adaptive receptive vocabulary assessment tool: An experimental study with Vietnamese EFL learners. *PASAA Journal*, 69, 533-560.
- Duong, T. T. H., Nguyen, T. T. H., & Hoang, T. M. (2024). Exploring teachers' perspectives on adaptive learning in undergraduate programs, Vietnam National University, Hanoi. *International Journal of TESOL & Education*, 4(3), 110-129. <https://doi.org/10.54855/ijte.24437>
- Educational Testing Service (ETS). (2023). *TOEFL® test updates and enhancements*. Retrieved from <https://www.ets.org>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- Lazendic, G., & Martin, A. J. (2018). Computer-adaptive testing: Implications for students' achievement, motivation, engagement, and subjective test experience. *Journal of Educational Psychology*, 110(1), [page numbers if available].
- Li, Z. (2020). Validity and fairness of the Duolingo English Test. *Language Testing*, 37(4), 627-635.
- Nguyen, A. Q. (2024). Innovation in Vietnamese higher education for society 5.0: Opportunities, challenges, and strategies. *International Journal of Research in Education and Science (IJRES)*, 10(4), 750-760. <https://doi.org/10.46328/ijres.3491>
- Wainer, H. (2000). *Computerized adaptive testing: A primer (2nd ed.)*. Lawrence Erlbaum Associates.
- Weiss, D. J. (2011). Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Sciences*, 2(1), 1-27.