BÀI THI TOEFL IBT PHIÊN BẢN 2026 TỪ GÓC NHÌN ĐÁNH GIÁ NGÔN NGỮ: THIẾT KẾ, THANG ĐIỂM VÀ HÀM Ý ĐỐI VỚI GIÁO DỤC NGOẠI NGỮ Ở VIỆT NAM

Trần Thị Thanh Bình
Học viện Công nghệ Bưu chính Viễn thông, Hà Nội
Email: binhttt@ptit.edu.vn

Tóm tắt: *Từ ngày 21/01/2026, bài thi TOEFL iBT được triển khai theo thiết kế mới: hai phần Đọc và Nghe áp dụng cách tiếp cận thích ứng đa chặng, các nhiệm vụ ở cả bốn kỹ năng được thay mới, và điểm được báo cáo trên thang 1–6 gắn với Khung tham chiếu ngôn ngữ chung châu Âu (CEFR). Bài viết phân tích những thay đổi này từ góc nhìn đánh giá ngôn ngữ nhằm trả lời ba câu hỏi: (1) phiên bản mới thay đổi những gì về cấu trúc, nhiệm vụ và báo cáo điểm; (2) các thay đổi đặt ra vấn đề gì khi xem xét theo khung lập luận hiệu lực; (3) chúng có hàm ý gì đối với dạy học, quy đổi chứng chỉ và đào tạo giáo viên ngoại ngữ tại Việt Nam. Nghiên cứu sử dụng phương pháp phân tích tài liệu trên ngữ liệu gồm tài liệu chính thức của ETS, một đánh giá độc lập và một số tài liệu hướng dẫn luyện thi, được mã hóa theo bốn chiều phân tích và sáu suy luận của lập luận hiệu lực. Kết quả cho thấy phiên bản mới (i) rút ngắn thời gian thi và thay đổi phương thức kiểm tra; (ii) chuyển trọng tâm sang các nhiệm vụ ngắn, độc lập, gần với giao tiếp trong đời sống đại học và không còn các nhiệm vụ tích hợp kỹ năng; (iii) giúp diễn giải điểm thuận lợi hơn nhờ thang điểm gắn CEFR. Tuy nhiên, bằng chứng độc lập về độ tin cậy chấm điểm, tính công bằng và hiệu lực dự báo vẫn còn hạn chế. Bài viết đề xuất một số điều chỉnh trong luyện thi, quy định công nhận chứng chỉ và đào tạo năng lực đánh giá cho giáo viên.*

Từ khóa: *đánh giá ngôn ngữ; TOEFL iBT 2026; kiểm tra thích ứng đa chặng; thang điểm 1–6; CEFR; hiệu ứng ngược.*

THE 2026 TOEFL IBT FROM A LANGUAGE ASSESSMENT PERSPECTIVE: DESIGN, SCORE SCALE, AND IMPLICATIONS FOR FOREIGN LANGUAGE EDUCATION IN VIETNAM

Abstract: *Since January 21, 2026, the TOEFL iBT has adopted a new design: multistage adaptive testing in Reading and Listening, revised tasks across all four skills, and scores reported on a 1–6 scale aligned with the Common European Framework of Reference for Languages (CEFR). This article examines these changes from a language assessment perspective, addressing three questions: (1) What has changed in test structure, tasks, and score reporting? (2) What issues do these changes raise within a validity argument framework? (3) What are their implications for teaching, certificate equivalency, and foreign language teacher education in Vietnam? The study employs document analysis of official ETS materials, an independent review, and selected test preparation resources, coded across four analytical dimensions and six validity argument inferences. The findings indicate that the new version (i) reduces testing time and changes the testing approach; (ii) shifts towards short, standalone tasks resembling communication in university life, while removing integrated-skills tasks; and (iii) facilitates score interpretation through CEFR alignment. However, independent evidence concerning scoring reliability, fairness, and predictive validity remains limited. The article proposes adjustments to test preparation, certificate recognition policies, and teacher assessment literacy development.*

Keywords: *language assessment; TOEFL iBT 2026; multistage adaptive testing; 1–6 score scale; CEFR; washback*

Nhận bài: 13/09/2026 *Phản biện: 13/10/2026* *Duyệt đăng: 17/10/2026*

I. ĐẶT VẤN ĐỀ

TOEFL iBT là một trong những bài thi năng lực tiếng Anh học thuật được sử dụng rộng rãi nhất thế giới. Theo ETS, kết quả bài thi được hơn 13.000 cơ sở giáo dục và tổ chức tại hơn 160 quốc gia chấp nhận. Tại Việt Nam, điểm TOEFL iBT được dùng để xét tuyển và học bổng du học, xét miễn học phần hoặc miễn thi ngoại ngữ, xác nhận chuẩn đầu ra của sinh viên, tuyển sinh sau đại học và đánh giá năng lực ngoại ngữ của giảng viên, viên chức. Trong nhiều năm, bài thi duy trì cấu trúc bốn kỹ năng với thang điểm 0–120 (mỗi kỹ năng 0–30). Từ ngày 21/01/2026, ETS áp dụng phiên bản điều chỉnh với ba nhóm thay đổi lớn: thiết kế thích ứng ở hai kỹ năng tiếp nhận, nhiều dạng nhiệm vụ mới và thang điểm 1–6 tương thích CEFR. Đối với một bài thi có hệ quả lớn (high-stakes), mỗi thay đổi về thiết kế đều kéo theo các câu hỏi về cái được đo, độ chính xác của phép đo và cách sử dụng kết quả. Các câu hỏi này liên quan trực tiếp đến những người làm công tác giảng dạy, kiểm tra đánh giá và quản lý đào tạo ngoại ngữ. Cho đến nay, thông tin về phiên bản mới chủ yếu xuất hiện dưới dạng tài liệu giới thiệu của nhà phát triển bài thi và các hướng dẫn luyện thi mang tính thương mại; phân tích học thuật theo khung đánh giá ngôn ngữ còn ít, nhất là trong bối cảnh Việt Nam. Bài viết góp phần lấp khoảng trống này bằng cách trả lời ba câu hỏi nghiên cứu:

RQ1. Phiên bản TOEFL iBT 2026 thay đổi những gì về cấu trúc, nhiệm vụ và báo cáo điểm so với phiên bản trước? RQ2. Những thay đổi đó đặt ra vấn đề gì về hiệu lực của bài thi khi xem xét theo khung lập luận hiệu lực? RQ3. Những thay đổi đó có hàm ý gì đối với dạy học và luyện thi, quy đổi chứng chỉ và đào tạo giáo viên ngoại ngữ ở Việt Nam?

II. CƠ SỞ LÝ LUẬN

2.1. Hiệu lực của bài thi và lập luận hiệu lực

Theo Messick (1996), hiệu lực không phải thuộc tính của bản thân bài thi mà là mức độ các cách diễn giải và sử dụng điểm số được bằng chứng và lý lẽ hỗ trợ. Kane (2013) cụ thể hóa quan điểm này thành lập luận dựa trên chuỗi suy luận từ hành vi làm bài đến quyết định sử dụng điểm. Chapelle, Enright và Jamieson (2008) áp dụng cách tiếp cận đó cho TOEFL với các suy luận: miêu tả miền sử dụng ngôn ngữ, đánh giá (chấm điểm), khái quát hóa, giải thích (cấu trúc đo), ngoại suy và sử dụng. Bài viết mượn sáu suy luận này làm khung để đọc các thay đổi của phiên bản 2026, vì khung cho phép đặt câu hỏi về từng mắt xích thay vì chỉ liệt kê các điểm mới.

2.2. Kiểm tra thích ứng đa chặng

Kiểm tra thích ứng điều chỉnh độ khó của nhiệm vụ theo biểu hiện của người dự thi để ước lượng năng lực chính xác hơn với ít câu hỏi hơn. Trong thiết kế đa chặng, bài thi được chia thành các mô-đun: người dự thi làm mô-đun đầu, rồi được chuyển sang mô-đun tiếp theo có độ khó phù hợp với kết quả vừa đạt. Cách làm này giúp nhà phát triển kiểm soát nội dung và chất lượng từng mô-đun tốt hơn so với thích ứng theo từng câu hỏi, đồng thời vẫn giữ được lợi thế về hiệu quả đo lường. Điểm của các dạng đề khác nhau được đưa về cùng một thang bằng phương pháp cân bằng điểm dựa trên lý thuyết ứng đáp câu hỏi (IRT).

2.3. CEFR và Khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam

CEFR (Council of Europe, 2001, 2020) mô tả năng lực ngôn ngữ ở sáu mức từ A1 đến C2. Khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam (Bộ Giáo dục và Đào tạo, 2014) được xây dựng trên cơ sở tham chiếu CEFR, với bậc 1 đến bậc 6 tương ứng A1 đến C2. Vì vậy, một bài thi quốc tế gắn thang điểm trực tiếp với CEFR có ý nghĩa thực tiễn đối với việc quy đổi và công nhận chứng chỉ trong nước.

2.4. Hiệu ứng ngược của bài thi

Hiệu ứng ngược (washback) là ảnh hưởng của bài thi đối với việc dạy và học (Alderson & Wall, 1993; Messick, 1996). Khi một bài thi có tầm quan trọng lớn thay đổi dạng nhiệm vụ, nội dung giảng dạy và luyện thi thường thay đổi theo. Do đó, phân tích cấu trúc nhiệm vụ là bước cần thiết để dự báo và định hướng hiệu ứng này.

III. PHƯƠNG PHÁP NGHIÊN CỨU

3.1. Thiết kế và ngữ liệu

Nghiên cứu có tính chất định tính, sử dụng phương pháp phân tích tài liệu. Ngữ liệu được thu thập đến ngày 29/10/2026 và chia thành ba nhóm theo mức độ tin cậy (Bảng 1).

Bảng 1. Ngữ liệu và vai trò trong phân tích

Nhóm	Nội dung	Vai trò
A. Tài liệu chính thức của ETS	Trang thông tin cập nhật tháng 01/2026; trang mô tả nội dung và cấu trúc bài thi; bản đặc tả và sơ đồ thiết kế bài thi 2026; bài giải thích thang điểm 1–6; tài liệu hỏi đáp dành cho giáo viên	Căn cứ chính cho các mô tả về thiết kế, thang điểm và cách tính điểm
B. Đánh giá độc lập	Báo cáo của một cơ sở đào tạo ngoại ngữ tham dự buổi giới thiệu của ETS tháng 11/2025	Cung cấp nhận định phản biện về giá trị của thay đổi
C. Tài liệu hướng dẫn luyện thi	Một số hướng dẫn thương mại mô tả định dạng và nhiệm vụ	Chỉ dùng để đối chiếu, bổ sung chi tiết khi phù hợp với nhóm A

3.2. Quy trình phân tích

Ngữ liệu được mã hóa theo bốn chiều: (a) cấu trúc và thời lượng; (b) dạng nhiệm vụ và ngữ cảnh sử dụng ngôn ngữ; (c) chấm điểm và báo cáo điểm; (d) mục đích sử dụng kết quả. Các mã sau đó được đối chiếu với sáu suy luận của lập luận hiệu lực (mục 2.1) để xác định những điểm đã có bằng chứng và những điểm còn bỏ ngỏ. Một thông tin chỉ được đưa vào phần kết quả khi có trong tài liệu nhóm A hoặc được ít nhất hai nguồn

độc lập xác nhận; thông tin còn lại được ghi rõ là chưa được xác nhận.

3.3. Hạn chế của phương pháp

Nghiên cứu chỉ dựa trên tài liệu công bố, không thu thập dữ liệu từ người dự thi và không có điều kiện tiếp cận dữ liệu kỹ thuật của ETS. Vì vậy, các kết luận về hiệu lực mang tính phân tích lập luận chứ không phải kiểm chứng thực nghiệm.

IV. KẾT QUẢ

4.1. Cấu trúc và phương thức triển khai (RQ1)

Theo ETS, từ tháng 01/2026 hai phần Đọc và Nghe được thiết kế theo dạng thích ứng đa chặng: giữa mỗi phần thi, bài thi điều chỉnh theo kết quả làm bài của thí sinh. Hai phần Nói và Viết là dạng tuyến tính, tức mọi thí sinh làm cùng một bộ nhiệm vụ trong một dạng đề. Theo bản đặc tả bài thi, thời gian làm bài tối thiểu là 1 giờ 23 phút và tối đa là 1 giờ 29 phút, chưa tính phần hướng dẫn; ETS khuyến nghị thí sinh dự trù khoảng hai giờ cho cả buổi thi. Do tính thích ứng, thời gian và số câu hỏi thực tế có thể khác nhau giữa các thí sinh. Một báo cáo tại buổi giới thiệu của ETS cho biết thứ tự các phần thi chuyển thành Đọc, Nghe, Viết, Nói. Các tài liệu hướng dẫn luyện thi nêu phần Nghe gồm 47 câu trong 29 phút và phần Nói gồm 11 câu trong 8 phút; thời lượng của hai phần còn lại chưa được xác nhận bằng tài liệu nhóm A nên bài viết không sử dụng.

4.2. Hệ thống nhiệm vụ (RQ1)

Bảng 2. Các dạng nhiệm vụ của TOEFL iBT từ tháng 01/2026

Kỹ năng	Nhiệm vụ	Mô tả khái quát
Đọc (thích ứng)	Complete the Words	Đoạn văn ngắn bị khuyết một phần chữ cái; thí sinh hoàn thành từ cho phù hợp ngữ cảnh
	Read in Daily Life	Thông báo, thư điện tử hoặc bài viết trực tuyến; tìm thông tin và suy luận
	Read an Academic Passage	Đoạn văn mang tính giáo trình đại học; câu hỏi về ý chính, từ vựng, chi tiết, suy luận
Nghe (thích ứng)	Listen and Choose a Response	Nghe một câu nói và chọn phản hồi phù hợp nhất
	Listen to a Conversation	Hội thoại ngắn có thành ngữ, cách nói thông tục; câu hỏi về thái độ, ý định, chi tiết
	Listen to an Announcement	Thông báo trong môi trường nhà trường, trường đại học
	Listen to an Academic Talk	Bài giảng học thuật ngắn
Viết	Build a Sentence	Sắp xếp các từ để tạo thành câu đúng ngữ pháp trong một hội thoại thân mật
	Write an Email	Viết thư điện tử đáp lại một tình huống thực tế
	Write for an Academic Discussion	Viết phản hồi trong một thảo luận trực tuyến ở lớp học
Nói	Listen and Repeat	Nghe một lần và nhắc lại chính xác câu nghe được (7 câu)
	Take an Interview	Trả lời các câu hỏi phỏng vấn về chủ đề quen thuộc (4 câu theo bản đặc tả)

Nguồn: tổng hợp từ ETS (2026) và các tài liệu nêu ở mục Tài liệu tham khảo.

Ba xu hướng nổi bật. Thứ nhất, xuất hiện các nhiệm vụ vi mô ở mức từ và câu (Complete the Words, Build a Sentence, Listen and Repeat), đây là các dạng hoàn toàn mới so với phiên bản trước. Thứ hai, miền sử dụng ngôn ngữ được mở rộng từ môi trường học thuật thuần túy sang các tình huống đời sống đại học và sinh hoạt thường nhật như thông báo, thư điện tử, thảo luận nhóm và phỏng vấn; ETS cho biết nội dung được chọn để phản ánh cách sinh viên dùng tiếng Anh trong thảo luận nhóm và làm dự án, đồng thời giảm thiên lệch văn hóa. Thứ ba, các nhiệm vụ tích hợp đọc–nghe–nói/viết như trước đây không còn trong cấu trúc mới.

4.3. Thang điểm và báo cáo điểm (RQ1)

Với các bài thi từ ngày 21/01/2026, thí sinh nhận bốn điểm thành phần và một điểm tổng thể trên thang 1–6, bước nhảy 0,5; điểm tổng thể là trung bình cộng của bốn điểm thành phần. Trong hai năm chuyển tiếp (đến tháng 01/2028), phiếu

điểm còn hiển thị điểm tổng thể tương đương trên thang 0–120. Điểm 0–120 này là giá trị trung điểm ước tính của khoảng điểm cũ ứng với mức điểm mới, không phải kết quả của một lần chấm riêng theo thang cũ. ETS giải thích rằng thang 1–6 được xây dựng với việc đối chiếu CEFR là một phần của quá trình phát triển thang, nhằm khắc phục tình trạng ngưỡng điểm tối thiểu để đạt cùng một mức CEFR khác nhau giữa các kỹ năng trong thang 0–30. Ở thang mới, cùng một giá trị điểm ở mọi kỹ năng và ở điểm tổng thể ứng với cùng một mức CEFR. Bảng 3 trình bày quan hệ tham chiếu giữa điểm, mức CEFR và bậc của Khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam.

Bảng 3. Điểm 1–6, mức CEFR và bậc trong Khung 6 bậc của Việt Nam

Điểm (thang 1–6)	Mức CEFR	Bậc (Khung 6 bậc)
6,0	C2	Bậc 6
5,0–5,5	C1	Bậc 5
4,0–4,5	B2	Bậc 4
3,0–3,5	B1	Bậc 3
2,0–2,5	A2	Bậc 2
1,0–1,5	A1	Bậc 1

Nguồn: tổng hợp từ ETS (2026), các tài liệu hướng dẫn và Bộ Giáo dục và Đào tạo (2014); việc đối chiếu bậc dựa trên sự tương ứng A1–C2 với bậc 1–6.

V. THẢO LUẬN

5.1. Các thay đổi nhìn từ lập luận hiệu lực (RQ2)

Về mặt tích cực, tính thích ứng giúp rút ngắn bài thi mà vẫn có cơ sở để duy trì độ chính xác, và việc gắn thang điểm với CEFR làm cho suy luận sử dụng rõ ràng hơn, nhất là với nơi sử dụng điểm có khung tham chiếu tương thích như Việt Nam. Về mặt thận trọng, các suy luận đánh giá, giải thích và ngoại suy hiện chủ yếu dựa trên tuyên bố của nhà phát triển. Báo cáo của Sabancı University School of Languages (2025) nhận xét rằng chưa rõ phiên bản mới có cải thiện trải nghiệm làm bài hoặc cung cấp chỉ báo tốt hơn về thành công học thuật hay không, và đề nghị đánh giá thêm về chất lượng đo lường, độ chính xác chấm điểm và tính công bằng sau khi bài thi được triển khai rộng rãi. Nhận định này phù hợp với yêu cầu của lập luận hiệu lực: tuyên bố về tính xác thực của nhiệm vụ cần được hỗ trợ bằng bằng chứng độc lập, không chỉ bằng mô tả thiết kế.

5.2. Hàm ý đối với dạy học và luyện thi (RQ3)

Thứ nhất, nội dung luyện thi cần chuyển từ "mẹo làm bài" sang phát triển năng lực sử dụng ngôn ngữ. Các nhiệm vụ như Complete the Words, Build a Sentence hay Listen and Repeat đòi hỏi nền tảng về từ vựng, hình thái từ, cú pháp và ngữ âm ở mức xử lý tự động, khó tối ưu bằng thủ thuật. Thứ hai, cần tăng cường các hoạt động giao tiếp có tính tương tác như phỏng vấn, trao đổi qua thư điện tử và thảo luận trực tuyến, trong đó có luyện phản hồi nhanh khi không có thời gian chuẩn bị (Take an Interview không có thời gian chuẩn bị theo mô tả của các tài liệu hướng dẫn). Thứ ba, do độ khó của chặng sau phụ thuộc vào kết quả chặng trước, người học cần được làm quen với hình thức thi trên máy tính có tính thích ứng và được hướng dẫn duy trì sự tập trung ngay từ đầu mỗi phần thi. Thứ tư, tài liệu cũ vẫn có giá trị cho việc phát triển năng lực ngôn ngữ chung, nhưng không còn phản ánh đầy đủ dạng nhiệm vụ và cấu trúc của bài thi mới; giáo viên nên ưu tiên tài liệu do ETS công bố hoặc phù hợp với phiên bản 2026.

5.3. Hàm ý đối với quy đổi chứng chỉ và quản lý đào tạo (RQ3)

Các cơ sở giáo dục cần rà soát quy định nội bộ về chuẩn đầu ra, điều kiện miễn thi ngoại ngữ và tiêu chuẩn tuyển sinh sau đại học để bảo đảm tương thích với phiếu điểm mới. Trong giai đoạn đến tháng 01/2028, nên chấp nhận cả hai thang điểm và có hướng dẫn thống nhất về cách đọc phiếu điểm, trong đó phân biệt kết quả của bài thi trước và sau ngày 21/01/2026. Điểm 0–120 trên phiếu điểm mới là giá trị ước tính theo trung điểm nên không nên dùng để so sánh từng điểm với các ngưỡng cũ. Về dài hạn, việc quy định ngưỡng theo bậc của Khung 6 bậc (hoặc mức CEFR tương ứng) sẽ ổn định hơn so với gắn với thang điểm riêng của từng chứng chỉ, và các cơ quan quản lý nên theo dõi các cập nhật của Bộ Giáo dục và Đào tạo về bảng quy đổi giữa chứng chỉ quốc tế và Khung 6 bậc.

5.4. Hàm ý đối với đào tạo giáo viên ngoại ngữ (RQ3)

Phiên bản mới là cơ hội để tăng cường năng lực đánh giá ngôn ngữ (language assessment literacy) của giáo viên: hiểu cấu trúc nhiệm vụ, hiểu ý nghĩa của thang điểm gắn CEFR, phân tích hiệu ứng ngược và thiết kế hoạt động kiểm tra đánh giá phù hợp trên lớp. Giáo viên cũng cần được chuẩn bị về mặt sư phạm để hỗ trợ người học có điều kiện tiếp cận công nghệ và thiết bị chưa đồng đều.

5.5. Hạn chế và hướng nghiên cứu tiếp theo

Bài viết có hai hạn chế chính: dựa trên tài liệu do chính nhà phát triển bài thi công bố cùng tài liệu thứ cấp, và chưa có dữ liệu từ người dự thi Việt Nam. Các nghiên cứu tiếp theo có thể (i) khảo sát trải nghiệm và kết quả của thí sinh Việt Nam; (ii) so sánh nội dung bài thi với nhu cầu giao tiếp học thuật thực tế của sinh viên; (iii) nghiên cứu hiệu ứng ngược đối với các lớp luyện thi; (iv) phân tích độ tin cậy và tính công bằng của chấm điểm khi ETS công bố thêm dữ liệu kỹ thuật.

VI. KẾT LUẬN

Phiên bản TOEFL iBT từ tháng 01/2026 là một lần điều chỉnh có hệ thống về thiết kế, nhiệm vụ và báo cáo điểm. Xét theo lập luận hiệu lực, bài thi mới tăng sức mạnh ở suy luận sử dụng nhờ thang điểm gắn CEFR và thuận lợi cho Việt Nam do Khung 6 bậc được xây dựng trên CEFR; đồng thời đặt ra nhiều câu hỏi mở ở các suy luận đánh giá, giải thích và ngoại suy, đòi hỏi bằng chứng độc lập. Đối với giáo dục ngoại ngữ ở Việt Nam, kết quả phân tích gợi ý ba hướng điều chỉnh: định hướng luyện thi theo phát triển năng lực sử dụng ngôn ngữ, cập nhật quy định công nhận chứng chỉ theo bậc năng lực và tăng cường năng lực đánh giá ngôn ngữ của giáo viên.

TÀI LIỆU THAM KHẢO

Alderson, J. C., & Wall, D. (1993). Does washback exist? Applied Linguistics, 14(2), 115–129.

Bachman, L. F., & Palmer, A. S. (2010). Language assessment in practice. Oxford University Press.

Bộ Giáo dục và Đào tạo (2014). Thông tư số 01/2014/TT-BGDĐT ngày 24/01/2014 ban hành Khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.) (2008). Building a validity argument for the Test of English as a Foreign Language. Routledge.

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge University Press.

Council of Europe (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume. Council of Europe Publishing.

Educational Testing Service (2026). TOEFL iBT Test: 2026 Update – Test Blueprint and Specifications Document. https://www.ets.org/content/dam/ets-org/pdfs/toefl/toefl-ibt-test-specifications-2026.pdf (truy cập 29/10/2026).

Educational Testing Service (2026). TOEFL iBT upcoming updates, January 2026. https://www.ets.org/toefl/test-takers/ibt/upcoming-updates-jan-2026.html (truy cập 29/10/2026).

Educational Testing Service (2026). Inside the TOEFL iBT test: Why the 1–6 score scale? https://www.ets.org/toefl/institutions/blog/why-1-to-6-score-scale (truy cập 29/10/2026).

Educational Testing Service (2026). TOEFL iBT test content, structure and administration. https://www.ets.org/toefl/score-users/about/content (truy cập 29/10/2026).

Educational Testing Service (2026). TOEFL iBT teacher FAQ. https://www.pt.ets.org/content/dam/ets-org/pt_pt/pdfs/toefl/teacher-faq.pdf (truy cập 29/10/2026).

Kane, M. T. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 1–73.

Messick, S. (1996). Validity and washback in language testing. Language Testing, 13(3), 241–256.

Sabancı University School of Languages (2025). 2026 TOEFL iBT: Key changes and considerations. https://sl.sabanciuniv.edu/en/node/795 (truy cập 29/10/2026).

Study.com Editorial Team (2026). What changed in the new TOEFL iBT exam by January 2026? https://study.com/resources/all-toefl-test-changes.html (truy cập 29/10/2026).