



Review Article

A Review on Vocabulary Tests of High Frequency English Words

Bui Thi Kim Phuong*

Hanoi University of Science and Technology, 1 Dai Co Viet, Hai Ba Trung, Hanoi, Vietnam

Received 02 August 2022

Revised 16 September 2022; Accepted 19 September 2022

Abstract: Vocabulary assessment plays a critical role in language education for EFL learners, educators and researchers. The article introduces and analyzes four English receptive vocabulary tests of high frequency words, including Vocabulary Levels Test, Vocabulary Size Test, Computer Adaptive Test of Size and Strength and New General Service List Test. Based on the review of the four tests, it provides practical suggestions for choosing an appropriate test to assess the vocabulary knowledge of EFL learners in Vietnam as well as implications for building new vocabulary tests to meet the demand of English language teaching and learning practices as well as the research needs in the field of language education in Vietnam.

Keywords: Receptive vocabulary, vocabulary tests, high frequency words, language education.

* Corresponding author.

E-mail address: phuong.buithikim@hust.edu.vn

<https://doi.org/10.25073/2588-1159/vnuer.4702>

Tổng quan về trắc nghiệm từ vựng tiếng Anh thông dụng

Bùi Thị Kim Phụng*

Đại học Bách Khoa Hà Nội, 1 Đại Cồ Việt, Hai Bà Trưng, Hà Nội, Việt Nam

Nhận ngày 02 tháng 8 năm 2022

Chỉnh sửa ngày 16 tháng 9 năm 2022; Chấp nhận đăng ngày 19 tháng 9 năm 2022

Tóm tắt: Kiểm tra từ vựng đóng vai trò quan trọng trong đào tạo ngôn ngữ với cả người học, và các nhà sư phạm cũng như các nhà nghiên cứu. Bài viết giới thiệu và phân tích bốn bài trắc nghiệm từ vựng tiếp nhận dùng để kiểm tra từ vựng tiếng Anh thông dụng, bao gồm bài trắc nghiệm cấp độ từ vựng, bài trắc nghiệm độ rộng từ vựng, bài trắc nghiệm thích ứng bằng máy tính về độ rộng và mức độ hiểu biết từ vựng và bài trắc nghiệm từ vựng thông dụng. Từ đó, bài viết đưa ra những gợi ý về việc lựa chọn một bài trắc nghiệm phù hợp để đánh giá năng lực từ vựng của người học tiếng Anh ở Việt Nam cũng như các đề xuất về việc xây dựng một bài trắc nghiệm đáp ứng nhu cầu của việc dạy và học tiếng Anh cũng như nhu cầu nghiên cứu trong lĩnh vực giáo dục ngôn ngữ tại Việt Nam.

Từ khóa: Từ vựng tiếp nhận, kiểm tra từ vựng, từ thông dụng, đào tạo ngôn ngữ.

1. Mở đầu

Kiểm tra kiến thức từ vựng có ý nghĩa quan trọng trong đào tạo ngôn ngữ với cả người dạy và người học. Với người học, việc kiểm tra kiến thức từ vựng giúp xác định trình độ thông thạo ngôn ngữ của người học vì kiến thức từ vựng đóng vai trò nền móng cho tất cả các hoạt động sử dụng ngôn ngữ [1]. Lĩnh hội được một lượng kiến thức từ vựng là một trong những điều kiện tiên quyết quan trọng để học ngôn ngữ thành công. Với người dạy, việc có những ước tính đáng tin cậy về kiến thức từ vựng cho phép giáo viên cung cấp tài liệu phù hợp với nhu cầu của người học, đánh giá hiệu quả của quá trình học và đặt ra các mục tiêu phù hợp để người học có thể phát triển kỹ năng và năng lực ngôn ngữ của mình [2].

Đối với mục đích nghiên cứu, kiến thức từ vựng trở thành một yếu tố dự báo mạnh mẽ về trình độ ngôn ngữ của người học và thậm chí cả thành tích học tập của họ [3]. Chiều hướng

ngược lại, năng lực từ vựng của người học có xu hướng cải thiện khi trình độ ngôn ngữ của họ phát triển [4], hay quá trình áp dụng bốn kỹ năng ngôn ngữ là đọc, nghe, nói và viết trong giao tiếp hỗ trợ việc thu nhận các từ mới học vào bộ nhớ [5]. Ngoài ra, các bài trắc nghiệm từ vựng có thể được sử dụng để để đánh giá tác động của trải nghiệm học tập đối với quá trình phát triển từ vựng cũng như để đo lường mức độ phát triển từ vựng [6].

Trong bối cảnh dạy và học ngoại ngữ Việt Nam, từ vựng luôn có được xem trọng trong chương trình giảng dạy tiếng Anh ở Việt Nam. Trong Chương trình giáo dục phổ thông môn Tiếng Anh được ban hành kèm theo Thông tư số 32/2018/TT-BGDĐT ngày 26 tháng 12 năm 2018 của Bộ trưởng Bộ Giáo dục và Đào tạo, học sinh sau khi hoàn thành chương trình phổ thông, cần có số lượng từ vựng là khoảng 2500 từ. Tuy nhiên, theo kết quả của một số lượng không nhiều các nghiên cứu gần đây đánh giá từ vựng của người học tiếng Anh của Việt Nam, học sinh phổ thông và sinh viên đại học có lượng từ vựng rất hạn chế, không đạt được yêu cầu về lượng từ này [7]. Bài viết thể hiện nỗ lực

* Tác giả liên hệ.

Địa chỉ email: phuong.buithikim@hust.edu.vn

<https://doi.org/10.25073/2588-1159/vnuer.4702>

của tác giả để góp phần cải thiện tình hình bằng cách tóm lược các bài trắc nghiệm từ vựng đang được nghiên cứu và sử dụng để đưa ra các đề xuất về việc lựa chọn cũng như xây dựng và phát triển các bài trắc nghiệm từ vựng phục vụ các mục đích khác nhau của người học, người dạy cũng như các nhà nghiên cứu trong lĩnh vực đào tạo ngôn ngữ ở Việt Nam.

2. Kiểm tra từ vựng tiếng Anh

Với những người ngoài lĩnh vực nghiên cứu ngôn ngữ, việc đánh giá xem một ai đó có biết hoặc không biết một từ hay nhiều từ trong ngôn ngữ khác là không hề khó khăn [8]. Cách đơn giản được nghĩ đến là đưa cho người đó một từ trong ngôn ngữ mẹ đẻ (first language) và yêu cầu từ tương đương trong ngôn ngữ khác hay còn gọi là ngôn ngữ đích (target language). Nếu người này đưa ra một từ chính xác, có thể đánh giá người đó biết ngôn ngữ đích. Tuy nhiên, trên thực tế việc kiểm tra đánh giá từ vựng không hề đơn giản mà đòi hỏi việc xem xét nhiều khía cạnh của từ vựng [9].

Một số tác giả đã đưa ra một cách tiếp cận phát triển đối với chiều sâu của kiến thức từ [10, 11]. Đối với các tác giả này, kiến thức từ chuyên sâu bao gồm từ việc nhận biết đơn thuần với từ đã gặp trước đó, đến có thể sử dụng từ một cách hiệu quả và theo cách phù hợp với ngữ cảnh. Một trong những cách đánh giá được biến đổi nhiều nhất về độ sâu của kiến thức từ vựng trong cách tiếp cận phát triển này là Thang kiến thức từ vựng (Vocabulary Knowledge Scale) của Paribakht và Welshe [10], được trình bày trong Bảng 1.

Bảng 1. Thang đo kiến thức từ vựng

I.	<i>Tôi không nhớ đã từ nhìn thấy từ này trước đây.</i>
II.	<i>Tôi đã nhìn thấy từ này trước đây, nhưng tôi không biết nghĩa của từ.</i>
III.	<i>Tôi đã nhìn thấy từ này trước đây, và tôi nghĩ nghĩa của từ đó là _____. (từ đồng nghĩa hoặc dịch nghĩa).</i>
IV.	<i>Tôi biết từ này, nghĩa của từ là _____. (từ đồng nghĩa hoặc dịch nghĩa).</i>
V.	<i>Tôi có thể sử dụng từ này trong câu _____. (viết câu).</i>

Các nhà ngôn ngữ học khác đã khái niệm hóa chiều sâu của kiến thức từ vựng không phải theo hướng tiếp cận phát triển, mà thay vào đó là hướng tiếp cận theo khía cạnh, cụ thể là người học có thể được đánh giá là biết một từ ở mức độ thành thạo một số khía cạnh khác nhau liên quan đến nó [12, 13]. Nation đã xác định một khung đánh giá từ vựng bao gồm ba khía cạnh chính, mỗi khía cạnh bao gồm ba khía cạnh nhỏ: i) Hình thức từ (các khía cạnh nhỏ: dạng nói, dạng viết và các thành tố của từ); ii) Nghĩa của từ (các khía cạnh nhỏ: hình thức và ý nghĩa, khái niệm và các tham chiếu, và các liên kết); và iii) Sử dụng từ (các khía cạnh nhỏ: chức năng ngữ pháp, kết hợp từ và các ràng buộc khi sử dụng) [12].

Dù việc kiểm tra từ vựng có các cách tiếp cận khác nhau, không thể phủ nhận rằng các bài trắc nghiệm từ vựng tiếp nhận dạng viết là phổ biến nhất, thường được tiến hành với khía cạnh cơ bản nhất và quan trọng nhất của kiến thức từ vựng, đó là mối quan hệ giữa dạng từ (form) và nghĩa (meaning), khía cạnh này làm nền móng để tiến hành việc học tập và lĩnh hội các khía cạnh khác của từ vựng [14]. Hiện nay, nhiều nghiên cứu về kiểm tra từ vựng của người học tiếng Anh tại Việt Nam đều sử dụng các bài trắc nghiệm từ vựng tiếp nhận dạng viết để xác định năng lực từ vựng của thí sinh, có thể kể đến các nghiên cứu kiểm tra từ vựng của sinh viên đại học của Le và Nation [15], Nguyen và Webb [16] và Dang [17] và các nghiên cứu kiểm tra từ vựng của học sinh phổ thông của Vu và Nguyen [18] và Nguyen [19].

3. Các bài trắc nghiệm từ vựng tiếng Anh

Từ vựng trở lại như một lĩnh vực được quan tâm trong những năm 1980-1990 [20], kể từ đó cả lĩnh vực giảng dạy và nghiên cứu chứng kiến sự gia tăng của các bài trắc nghiệm từ vựng. Trong khuôn khổ bài viết này, với mục đích tổng quan các bài trắc nghiệm từ vựng thông dụng tiếp nhận hướng tới đối tượng người học tiếng Anh ở Việt Nam, bốn bài trắc nghiệm sẽ được tổng lược, bao gồm: i) Bài trắc nghiệm cấp độ từ vựng ii) Bài trắc nghiệm độ rộng từ vựng; iii) Bài trắc nghiệm thích ứng bằng máy

tính về độ rộng và mức độ hiểu biết từ vựng; và iv) Bài trắc nghiệm từ vựng thông dụng. Các thông tin cũng như những ưu nhược sẽ được thảo luận để từ đó đi đến những đề xuất về việc lựa chọn đề kiểm tra từ vựng phù hợp cho đối tượng người học tiếng Anh tại Việt Nam cũng như việc xây dựng và phát triển các bài trắc nghiệm từ vựng trong thời gian tới.

3.1. Bài trắc nghiệm cấp độ từ vựng

Bài trắc nghiệm cấp độ từ vựng được công bố từ 1983 và được đón nhận nhất trong việc giảng dạy [17]. Bài trắc nghiệm cấp độ từ vựng là một công cụ để đo lường từ vựng dạng viết của thí sinh trong năm nhóm từ vựng. Các nhóm 2.000 từ đầu tiên lấy từ danh sách từ thông dụng General Service List (GSL) của West (1953), và các nhóm 3.000, 5.000 và 10.000 từ lấy các từ thông dụng trong danh sách của Thorndike và Lorge (1944) và Kucera và Francis (1967) [21]. Trên trang web <http://www.lex tutor.ca/tests>, có thể truy cập các bài trắc nghiệm cấp độ từ vựng, đó là bài trắc nghiệm cấp độ từ vựng 2k-10k của Nation (1990) [22], bài trắc nghiệm cấp độ từ vựng 1k của Nation (2001) [23], bài trắc nghiệm cấp độ từ vựng phiên bản 2 của Schmitt và cộng sự (2001) [24] và Bài trắc nghiệm cấp độ từ vựng cập nhật của Webb, Sasao và Ballance (2017) [25].

Mục đích chính của bài trắc nghiệm là để chẩn đoán mức độ thông thạo từ vựng thông dụng của người học để chỉ định tài liệu học tập phù hợp. Mức độ thành thạo được xác định trên cơ sở trả lời đúng 29 câu hỏi hoặc nhiều hơn cho 30 câu hỏi tương ứng, vì cần có độ bao phủ 98% để đọc hiểu các tài liệu viết [12]. Là một bài trắc nghiệm chẩn đoán, Bài trắc nghiệm cấp độ từ vựng được chứng minh là rất hữu ích trong việc hướng sự quan tâm đến mức độ từ vựng tương ứng của người dạy và người học. Các câu hỏi của bài trắc nghiệm cấp độ từ vựng được thiết kế với cùng một định dạng ghép nối, trong đó người học phải ghép ba trong số sáu từ ở bên trái với các nghĩa được liệt kê ở bên phải.

Sau khi hoàn thành câu trả lời, trang web sẽ cho điểm về tỷ lệ phần trăm câu trả lời đúng trong số các từ được kiểm tra và đưa ra gợi ý về

các nhóm vấn đề nếu câu trả lời sai ở các cấp độ.

Các phiên bản khác nhau đã được áp dụng rộng rãi trong cả đánh giá và nghiên cứu, và khá nhiều bài báo đã được xuất bản để xác định bài trắc nghiệm. Read thấy Bài trắc nghiệm cấp độ từ vựng phiên bản 1983 đáng tin cậy [26]. Schmitt và cộng sự kết hợp bốn phiên bản gốc thành hai phiên bản, mỗi phiên bản có 10 cụm trong năm phần [24]. Hai phiên bản này được tiến hành xác định với tổng cộng 801 thí sinh thuộc 13 nhóm ở năm quốc gia. Các chỉ số độ tin cậy (Cronbach's alpha) cho tất cả các phần tương ứng các cấp độ đều cao. Các cuộc phỏng vấn với thí sinh cũng chỉ ra rằng thí sinh đón nhận bài trắc nghiệm và bài trắc nghiệm có thể phản ánh kiến thức từ vựng cơ bản của họ. Ngoài ra, bài trắc nghiệm cũng được chỉ ra là phân biệt tốt giữa người học giỏi hơn và người học yếu hơn và do đó bài trắc nghiệm có thể được sử dụng như bài trắc nghiệm xếp lớp cũng như xếp người học nhanh chóng vào các nhóm năng lực khác nhau dựa trên kiến thức từ vựng.

Tuy nhiên, định dạng câu hỏi của bài trắc nghiệm cấp độ từ vựng trong các phiên bản trước đó được các học giả cho rằng có một số khuyết điểm: a) thiếu tính độc lập của các câu hỏi, b) sự không chính xác tương đối của định dạng khi so với một câu hỏi có bốn lựa chọn tiêu chuẩn, c) học sinh khó hiểu định dạng và d) khó điều chỉnh các bài trắc nghiệm với các hình thức kiểm tra khác hoặc kho ngữ liệu nguồn [27, 28]. Đây cũng là lý do định dạng câu hỏi riêng lẻ cho mỗi từ với bốn lựa chọn được sử dụng trong phiên bản bài trắc nghiệm cấp độ từ vựng mới năm 2015.

Ngoài vấn đề về định dạng câu hỏi, bài trắc nghiệm cấp độ từ vựng cũng thiếu những bằng chứng xác định như Schmitt và cộng sự [29] đã chỉ ra: bài trắc nghiệm cấp độ từ vựng phiên bản 1983 chỉ có một đoạn ngắn đề cập đến việc tiến hành kiểm tra thử và kết luận bài trắc nghiệm không nên được sử dụng với những người học có tiếng mẹ đẻ là ngôn ngữ Latin; bài trắc nghiệm cấp độ từ vựng phiên bản 2001 có đề cập một nghiên cứu với 801 người có ngôn ngữ mẹ đẻ khác nhau và một số bằng chứng xác thực, nhưng không bàn về cách thức bài trắc

nhịệm nên được sử dụng với những người học khác nhau hoặc trong các ngữ cảnh khác nhau; bài báo về bài trắc nghiệm cấp độ từ vựng phiên bản 2015 cũng không tập trung vào vấn đề xác trị. Thêm vào đó, bài trắc nghiệm cấp độ từ vựng cũng ko có thông tin về ý nghĩa của điểm kiểm tra liên quan đến những gì người học thực sự có thể làm với các từ được kiểm tra (ngoài việc trả lời chúng trong bài trắc nghiệm) trong khi thông tin này rất có giá trị với người học và đáp ứng một mức độ xác thực nhất định của bài trắc nghiệm.

3.2. Bài trắc nghiệm độ rộng từ vựng

Bài trắc nghiệm độ rộng từ vựng là một bài trắc nghiệm gần đây hơn về kích thước từ vựng của Nation và Beglar [30]. Bài trắc nghiệm độ rộng từ vựng được thiết kế để đo lường kiến thức từ vựng tiếp thu ở dạng viết, ước tính tổng số lượng từ vựng của thí sinh. Đây là một bài trắc nghiệm trình độ nhưng ước tính cũng có thể được sử dụng để chẩn đoán sự tăng trưởng vốn từ vựng dài hạn và đặt mục tiêu học từ vựng mới. Bài trắc nghiệm độ rộng từ vựng có hai phiên bản, một phiên bản 14.000 chứa 140 câu hỏi trắc nghiệm và hai phiên bản 20.000 song song chứa 100 câu hỏi trắc nghiệm. Cả hai phiên bản kiểm tra giấy 14.000 và 20.000 đều có thể truy cập tại <http://www.victoria.ac.nz/lals/about/staff/paul-nation> và thí sinh có thể kiểm tra với máy tính khi truy cập tại <http://my.vocabularysize.com>. Một số phiên bản song ngữ của bài trắc nghiệm độ rộng từ vựng đã được phát triển bao gồm tiếng A Rập, tiếng Gujarati, tiếng Nga, tiếng Hàn, tiếng Việt, tiếng Quan thoại và tiếng Nhật. Trong các phiên bản song ngữ, các từ đích là tiếng Anh và các lựa chọn sử dụng tiếng mẹ đẻ của người học. Các phiên bản song ngữ được nhận định gia tăng giá trị của bài trắc nghiệm từ vựng tiếp nhận khi vẫn kiểm tra đánh giá kiến thức từ vựng tiếp nhận mà giảm thiểu các yếu tố nhiễu như ngữ pháp hay kỹ năng đọc hiểu tiếng Anh của thí sinh [15].

bài trắc nghiệm độ rộng từ vựng sử dụng định dạng câu hỏi trắc nghiệm với bốn phương án lựa chọn nghĩa của từ được kiểm tra. Bài trắc nghiệm đưa ra từ được kiểm tra bằng một câu đã được loại bỏ ngữ cảnh cùng với bốn phương

án thay thế ý nghĩa. Tổng điểm cần được nhân với 100 đối với phiên bản 14,000 và 200 đối với phiên bản 20,000 để tính tổng lượng từ vựng của thí sinh. Kết quả báo cáo tỷ lệ phần trăm kiến thức về từ của người học ở mỗi cấp độ (kiến thức một phần cũng được đánh giá ở đây) và tổng số lượng từ vựng ước tính và do đó bài trắc nghiệm có thể được sử dụng như một bài trắc nghiệm chẩn đoán độ rộng từ vựng tiếp thu. Le và Nation [15] và Karami [31] đề xuất rằng mọi cấp độ của bài trắc nghiệm nên được kiểm tra để tránh việc kích thước từ vựng bị đánh giá thấp đi nhiều so với thực tế và có được ước tính hợp lệ hơn về kích thước từ vựng của người học, mặc dù một số nghiên cứu khác ủng hộ rằng học sinh chỉ cần làm kiểm tra hai mức trên khả năng của họ để có độ chính xác ngày càng cao của kết quả bài trắc nghiệm độ rộng từ vựng [21].

Về vấn đề xác trị bài trắc nghiệm, bài báo đầu tiên không có thông tin xác thực, nhưng một nghiên cứu tiếp theo của Beglar [32] cung cấp bằng chứng hợp lệ cho bài trắc nghiệm độ rộng từ vựng bằng cách xác định độ giá trị cấu trúc về nội dung, cơ sở lý thuyết, cấu trúc, độ tin cậy từ khung xác trị của Messick và nghiên cứu khả năng đáp ứng và khả năng diễn giải. Tính đơn chiều của bài trắc nghiệm cũng được đánh giá với cách tiếp cận dựa vào Rasch. Có thể thấy rằng bài trắc nghiệm độ rộng từ vựng có thể phân biệt người học với các mức độ khác nhau của kiến thức từ vựng tiếp thu dạng viết và cung cấp đủ cấp độ từ để đo lường khả năng tiếp thu từ vựng của người học trong một khoảng thời gian dài. Riêng với phiên bản song ngữ tiếng Việt, thông tin xác trị được đưa ra với lượng thí sinh khá nhỏ, chỉ 62 sinh viên chuyên ngữ năm thứ 3 tại một trường đại học [15]. Schmitt và cộng sự [29], dựa vào thông tin tổng hợp về các phiên bản song ngữ, đã đưa ra đề xuất về việc tiến hành thêm các nghiên cứu cung cấp thêm bằng chứng xác trị hợp lệ với các phiên bản song ngữ, trong đó có phiên bản song ngữ tiếng Việt.

Ngoài ra, các nhà nghiên cứu cũng lưu ý việc suy luận thận trọng khi sử dụng các kết quả từ bài trắc nghiệm độ rộng từ vựng để đo kiến thức từ vựng của người học [29]. Vì bài trắc

nghiệm độ rộng từ vựng lấy mẫu mỗi 10 từ từ các nhóm 1.000 từ phổ biến nhất (lên đến nhóm thứ 14 hoặc 20 tùy thuộc vào phiên bản), “điểm của thí sinh cần được nhân với 100 để có tổng lượng từ vựng của họ” [12]. Ví dụ, điểm 30 trên 140 sẽ tạo ra ước tính kích thước của 3.000 họ từ đã biết. Có thể thấy suy luận này dễ hiểu và dễ tiến hành, tuy nhiên, có hai vấn đề cần lưu ý ở đây, đó là tính đại diện của 10 từ lấy mẫu trong đề kiểm tra và việc sử dụng định dạng câu hỏi trắc nghiệm 4 lựa chọn mang lại 25% khả năng thí sinh trả lời đúng câu hỏi nhưng không chắc chắn thí sinh thực sự biết từ kiểm tra. Để giảm sai số trong kiểm tra từ vựng của thí sinh, các tác giả đưa ra đề xuất về việc sử dụng các mô hình lý thuyết hỏi đáp IRT nhiều tham số hay áp dụng kiểm tra thích ứng với các ứng dụng công nghệ trong quá trình xây dựng các bài trắc nghiệm từ vựng.

3.3. Bài trắc nghiệm thích ứng bằng máy tính về độ rộng và mức độ hiểu biết từ vựng

Bài trắc nghiệm thích ứng bằng máy tính về độ rộng và mức độ hiểu biết từ vựng (Computer Adaptive Test of Size and Strength – CATSS, từ đây gọi là bài trắc nghiệm CATSS) hướng tới nhiều mục đích theo Schmitt và cộng sự [29] tổng kết: i) cung cấp cho các nhà nghiên cứu và giáo viên một hồ sơ chẩn đoán về mức độ phát triển ngôn ngữ từ vựng hiện tại của người học; ii) cung cấp một phương tiện đo lường sự phát triển cả về độ rộng và mức độ hiểu biết khi tiếp nhận từ vựng; iii) cho phép xếp lớp hiệu quả hơn trong các chương trình giảng dạy ngôn ngữ vì kiến thức từ vựng đã được chứng minh là có liên quan đến sự thành công trong việc đọc, viết và thông thạo ngôn ngữ nói chung cũng như thành tích học tập. Bài trắc nghiệm CATSS có thể truy cập tại <https://lexisite.com/catss/>.

Yếu tố độ rộng hay kích thước từ vựng của bài trắc nghiệm CATSS được phản ánh trong các mẫu từ từ năm nhóm: 1 - nhóm 2000 từ phổ biến nhất, 2 - nhóm 1000 từ phổ biến thứ ba, 3 - nhóm 1000 từ phổ biến thứ năm, 4 - nhóm 1000 từ phổ biến thứ mười và 5- nhóm từ vựng học thuật AWL (Academic Word List).

Yếu tố mức độ hiểu biết được phản ánh trong bốn mức độ hiểu biết cho mỗi từ, bắt đầu

với mức khó nhất đến mức dễ nhất tương ứng với các yêu cầu ở mức khó nhất - cung cấp từ với nghĩa cho trước đến yêu cầu ở mức dễ nhất – chọn đáp án chứa nghĩa đúng nhất trong các đáp án cho sẵn. Do đó, việc cho điểm của các mức độ hiểu biết khác nhau trong phiên bản bài trắc nghiệm CATSS phản ánh độ khó khác nhau này: 1 điểm cho câu trả lời đúng ở mức khó nhất, 0,75 điểm và 0,5 với hai mức tiếp theo và 0,25 điểm với mức dễ nhất.

So với các bài trắc nghiệm từ vựng được tóm lược trong bài viết này, bài trắc nghiệm CATSS là bài trắc nghiệm từ vựng duy nhất sử dụng hình thức thích ứng bằng máy tính. Với hình thức này, nếu thí sinh nhận được một câu hỏi về một từ với một mức độ hiểu biết nhất định, tùy câu trả lời của thí sinh, thí sinh sẽ nhận câu hỏi tiếp về từ khác hoặc về từ đó nhưng với các mức độ hiểu biết tương thích. Ví dụ: nếu thí sinh đã cung cấp từ có nghĩa cho trước (mức khó nhất), từ đó sẽ không xuất hiện lại trong ba mức còn lại. Khi không có phản hồi hoặc phản hồi không chính xác được đưa ra, máy tính sẽ giữ từ đó trong bộ nhớ của nó để kiểm tra từ đó ở mức dễ hơn, nhưng chỉ sau khi tất cả các từ khác ở cùng nhóm đã được kiểm tra ở mức đó. Một khi thí sinh trả lời đúng, từ đó sẽ không được kiểm tra lại với các mức thấp hơn.

Cần lưu ý rằng trong bài trắc nghiệm CATSS, từ vựng tiếp nhận chỉ là một nội dung kiểm tra đánh giá. Bài trắc nghiệm CATSS cung cấp một bức tranh toàn diện hơn cả về độ rộng và mức độ hiểu biết từ vựng của thí sinh khi bài trắc nghiệm CATSS không chỉ tính đến số lượng từ mà người dự thi biết mà còn tính đến mức độ hiểu biết của những từ này [33, 34], tuy nhiên không thể phủ nhận được một số thách thức bài trắc nghiệm CATSS phải đối mặt. Về định dạng của câu hỏi, có 4 dạng tương ứng với 4 mức độ hiểu biết từ vựng. Mặc dù hình thức thích ứng bằng máy tính được áp dụng, mỗi từ chỉ kiểm tra một mức độ hiểu biết, số lượng câu hỏi của bài trắc nghiệm là 140 câu có thể được xem là khá dài, cùng với định dạng khác nhau cho các câu hỏi có thể mang lại những khó khăn không nhỏ đối với thí sinh.

Về định dạng câu hỏi với mức độ hiểu biết cao nhất, thí sinh được yêu cầu cung cấp dạng từ, chính những học giả thiết kế bài trắc nghiệm CATSS cũng ghi nhận đó là một thách thức lớn vì câu trả lời của thí sinh đưa ra có sự khác biệt lớn phụ thuộc vào trình độ của từng thí sinh,

việc chuẩn bị đầy đủ đáp án để tiến hành chấm tự động trên máy tính đòi hỏi nỗ lực rất lớn của những nhà thiết kế cùng với những chuyên gia là người bản địa và những người có năng lực tương đương với thí sinh.

Bảng 2. Định dạng câu hỏi trong bài trắc nghiệm CATSS

Nội dung kiến thức	Tính điểm	Yêu cầu của câu hỏi	Ví dụ
Productive recall	1	Cung cấp từ có nghĩa cho trước	The task is to supply the L2 target word (melt). Turn into water _____
Receptive recall	0,75	Điền từ phù hợp vào một cụm từ hoặc một câu ngắn để giải thích nghĩa của từ cho trước	The task is to demonstrate understanding of the meaning of the L2 word (melt). When something melts, it turns into _____.
Productive recognition	0,5	Chọn đáp án đúng chứa từ từ có nghĩa cho trước trong số bốn phương án a-b-c-d.	The task is to choose the target word from four options (a b c d). Turn into water a. elect b. blame c. melt d. threaten
Receptive recognition	0,25	Chọn đáp án có giải thích đúng nhất nghĩa của từ cho trước trong bốn phương án 1-b-c-d.	The task is to choose the meaning of the target word from the four options provided. When something melts, it _____ a. chooses b. accuses c. makes threats d. turns into water

Ngoài ra, bài trắc nghiệm CATSS ban đầu sử dụng các từ được lấy mẫu không theo trình tự các nhóm từ - K1 + K2, K3, K5, K10 và danh sách từ học thuật (Academic word list). Do đó, điều này có thể hạn chế việc đánh giá tổng lượng từ vựng của người học [21]. Đây cũng là lý do mà bài trắc nghiệm CATSS mới bản 2019 [35] khắc phục hạn chế này bằng cách kiểm tra các từ từ 14 nhóm từ tuần tự theo cách tương tự như trong bài trắc nghiệm độ rộng từ vựng của Nation và Beglar [30].

Có thể thấy rằng bài trắc nghiệm CATSS có những ưu thế khi áp dụng hình thức trắc nghiệm

thích ứng bằng máy tính cùng với các bằng chứng xác thực đáng tin cậy, nhưng việc đáp ứng nhiều mục đích kiểm tra và việc đưa vào nhiều nội dung kiểm tra cùng nhiều định dạng câu hỏi có thể mang lại những ảnh hưởng không nhỏ đến trải nghiệm cũng như kết quả thể hiện của các thí sinh.

3.4. Bài trắc nghiệm từ vựng thông dụng

Bài trắc nghiệm từ vựng thông dụng (New General Service List Test) là một bài trắc nghiệm chẩn đoán về kiến thức tiếp nhận từ vựng tiếng Anh với danh sách từ thông dụng New General Service List (từ đây gọi là danh sách từ NGSL). Danh sách từ này được công bố

vào năm 2013 do nhóm tác giả Browne, Culligan và Phillips, có cùng mục đích như danh sách từ thông dụng General Service List của West (1953) được biết đến rộng rãi trước đây gồm các từ vựng cốt lõi có tần suất sử dụng cao cho người học tiếng Anh như ngôn ngữ thứ hai hay ngoại ngữ. Tuy nhiên, danh sách từ NGSL có tính cập nhật hơn với độ phủ hơn 92% hầu hết các văn bản tiếng Anh thông thường (mức cao nhất trong bất kỳ danh sách từ tiếng Anh thông dụng có nguồn gốc từ ngữ liệu cho đến nay) [36].

Bài trắc nghiệm có thể truy cập tại <http://www.newgeneralservicelist.org/> với 100 câu hỏi, 20 câu hỏi cho mỗi dải trong số năm dải từ khoảng 560 từ của danh sách từ NGSL. Việc chia thành 5 dải từ như vậy được giải thích để mang lại độ chính xác cao hơn trong việc phân loại người học ngoại ngữ tiếng Anh, từ đó đáp ứng nhu cầu về kế hoạch học tập phù hợp với các nhóm kết quả khác nhau.

Bài trắc nghiệm từ vựng thông dụng có những đặc điểm được xem là khắc phục các vấn đề gây khúc mắc của hai bài trắc nghiệm độ rộng từ vựng và bài trắc nghiệm cấp độ từ vựng. Bài trắc nghiệm từ vựng thông dụng bao gồm 100 câu hỏi, 20 câu hỏi được lựa chọn với mỗi dải từ để đại diện cho khoảng biến thiên và mức độ trung bình về độ khó của mỗi dải từ. Trong mỗi câu hỏi, từ đích được cung cấp và sau đó là một câu sử dụng từ trong ngữ cảnh. Tiếp theo là bốn lựa chọn câu trả lời trong đó có ba phương án nhiễu và câu trả lời đúng. Cách diễn giải kết quả của bài trắc nghiệm từ vựng thông dụng cung cấp thông tin về các cấp độ từ vựng như bài trắc nghiệm cấp độ từ vựng, từ đó mang lại những gợi ý hữu ích cho thí sinh trong việc tích lũy từ vựng sau bài trắc nghiệm. Ngoài phiên bản gốc tiếng Anh, Stoeckel và đồng sự đã phát triển phiên bản song ngữ tiếng Nhật và cung cấp các bằng chứng xác trị đáng tin cậy [37].

Xét trong bối cảnh dạy và học ngoại ngữ Việt Nam, bài trắc nghiệm từ vựng thông dụng có thể được xem là một lựa chọn phù hợp để đánh giá lượng từ vựng của người học, cũng như giúp xác định mục tiêu từ vựng trong kế hoạch dạy và học và lựa chọn các giải pháp dạy

và học phù hợp. Ngoài bài trắc nghiệm về danh sách từ thông dụng, dự án NGSL cũng cung cấp các công cụ hỗ trợ như danh sách từ vựng, website học từ hay các phần mềm cài đặt trên điện thoại mang lại nhiều cơ hội và trải nghiệm khác nhau để người học có thể tích lũy vốn từ vựng thông dụng như một nhiệm vụ khả thi trong quá trình phát triển năng lực ngôn ngữ tiếng Anh và rèn luyện các kỹ năng thực hành ngôn ngữ của mình.

Với các dự án phát triển công cụ kiểm tra đánh giá từ vựng cho người học Tiếng Anh tại Việt Nam, các nhà nghiên cứu có thể xem xét các gợi ý của nhóm tác giả bài trắc nghiệm từ vựng thông dụng như việc phát triển các phiên bản tương đương phục vụ cho việc tiến hành kiểm tra đánh giá thường xuyên và phiên bản song ngữ với các ngôn ngữ khác cũng như việc thực hiện các nghiên cứu xác trị cho Bài trắc nghiệm từ vựng thông dụng. Bên cạnh đó, đề xuất của Stoeckel và đồng sự [38] về việc áp dụng lý thuyết hồi đáp IRT để phát triển đề kiểm tra thích ứng cũng nên được xem xét nhằm mang lại những bài trắc nghiệm từ vựng với ít câu hỏi hơn, tiết kiệm thời gian hơn mà vẫn mang lại kết quả chính xác về năng lực thí sinh.

4. Kết luận

Từ việc tổng quan các nghiên cứu và các bài trắc nghiệm từ vựng tiếng Anh thông dụng đang được sử dụng có thể đi đến kết luận các bài trắc nghiệm từ vựng đều có giá trị thực tiễn và đang được sử dụng rộng rãi trong dạy và học tiếng Anh cũng như trong nghiên cứu. Việc lựa chọn một bài trắc nghiệm tùy thuộc vào mục đích sử dụng khác nhau của người học, người dạy và các nhà nghiên cứu, cũng như mức độ quen thuộc của người học với các dạng thức kiểm tra, từ đó mang lại những trải nghiệm tích cực cũng như kết quả thể hiện chính xác năng lực của thí sinh. Ngoài ra, qua việc tổng lược các bài trắc nghiệm từ vựng tiếp nhận tiếng Anh, tác giả cũng đề xuất một số vấn đề cần được quan tâm trong các dự án về kiểm tra từ vựng trong tương lai, cụ thể là các vấn đề về xác định mục đích kiểm tra rõ ràng, lấy mẫu từ kiểm tra phù hợp mục đích kiểm tra, xác định

định dạng câu hỏi phù hợp, phát triển phiên bản song ngữ, áp dụng công nghệ máy tính trong thiết kế cũng như trong thực hiện kiểm tra, và tiến hành xác trị nghiêm túc.

Tài liệu tham khảo

- [1] N. Schmitt, T. Cobb, M. Horst, D. Schmitt, How much Vocabulary is Needed to use English? Replication of van Zeeland and Schmitt (2012), Nation (2006) and Cobb (2007), *Language Teaching*, Vol. 50, No. 2, 2017, pp. 212-226.
- [2] I. S. P. Nation, How many High Frequency Words are there in English, *Language, Learning and Literature: Studies Presented to Hakan Ringbom*, Abo Akademi University, Abo: English Department Publications, Vol. 4, 2001, pp. 167-181.
- [3] L. H. Lin, B. Morrison, The Impact of the Medium of Instruction in Hong Kong Secondary Schools on Tertiary Students' Vocabulary, *Journal of English for Academic Purposes*, Vol. 9, No. 4, 2010, pp. 255-266.
- [4] A. Zareva, P. Schwanenflugel, Y. Nikolova, Relationship between Lexical Competence and Language Proficiency: Variable Sensitivity, *Studies in Second Language Acquisition*, Vol. 27, No. 4, 2005, pp. 567-595.
- [5] B. Laufer, Lexical Thresholds for Reading Comprehension: What they are and How they can be Used for Teaching Purposes, *Tesol Quarterly*, Vol. 47, No. 4, 2013, pp. 867-872.
- [6] T. Stoeckel, P. Bennett, A Test of the New General Service List, *Vocabulary Learning and Instruction*, Vol. 4, No. 1, 2015, pp. 1-8.
- [7] D. V. Vu, E. Peters, Vocabulary in English Language Learning, Teaching, and Testing in Vietnam: A review, *Education Sciences*, Vol. 11, No. 9, 2021, p. 563.
- [8] J. Read, *Assessing Vocabulary*, Cambridge University Press, 2000.
- [9] N. Schmitt, Size and Depth of Vocabulary Knowledge: What the Research Shows, *Language Learning*, Vol. 64, No. 4, 2014, pp. 913-951.
- [10] T. S. Paribakht, M. Wesche, Vocabulary Enhancement Activities and Reading for Meaning in Second Language Vocabulary Acquisition, *Second Language vVocabulary Acquisition: A Rationale for Pedagogy*, Vol. 55, No. 4, 1997, pp. 174-200.
- [11] N. Schmitt, C.B. Zimmerman, Derivative word forms: What do Learners Know?, *TESOL Quarterly*, Vol. 36, No. 2, 2002, pp. 145-171.
- [12] I. S. P. Nation, *Learning Vocabulary in Another Language*, Cambridge University Press, 2013.
- [13] N. Schmitt, *Researching vocabulary: A Vocabulary Research Manual*, Basingstoke: Palgrave Macmillan, 2010.
- [14] S. A. Webb, A. C. S. Chang, Second Language Vocabulary Growth, *RELC Journal*, Vol. 43, No. 1, 2012, pp. 113-126.
- [15] T. C. N. Le, P. Nation, A Bilingual Vocabulary Size Test of English for Vietnamese Learners, *RELC Journal*, Vol. 42, No. 1, 2011, pp. 86-99.
- [16] T. M. H. Nguyen, S. Webb, Examining Second Language Receptive Knowledge of Collocation and Factors that Affect Learning, *Language Teaching Research*, Vol. 21, No. 3, 2017, pp. 298-320.
- [17] T. N. Y. Dang, Vietnamese Non-English Major EFL University Students' Receptive Knowledge of the Most Frequent English Words, *VNU Journal of Foreign Studies*, Vol. 36, No. 3, 2020, pp. 1-11.
- [18] D. V. Vu, N. C. Nguyen, An Assessment of Vocabulary Knowledge of Vietnamese EFL Learners, presented at The 20th English in Southeast Asia Conference, National Institute of Education, Nanyang Technological University, Singapore, 2019.
- [19] C. D. Nguyen, Lexical Features of Reading Passages in English-language Textbooks for Vietnamese High-school Students: Do they Foster both Content and Vocabulary Gain?, *RELC Journal*, Vol. 52, No. 3, 2021, pp. 509-522.
- [20] P. M. Meara, *Vocabulary in a Second Language (Vol. 2. Specialized Bibliography 4)*, London: Centre for Information on Language Teaching and Research (CILT), 1987.
- [21] S. McLean, B. Kramer, The Creation of a New Vocabulary Levels Test, *Shiken*, Vol. 19, No. 2, 2015, pp. 1-11.
- [22] I. S. P. Nation, *Teaching and Learning Vocabulary*, Boston, MA: Heinle and Heinle, 1990.
- [23] I. S. P. Nation, *Learning Vocabulary in Another Language*, Cambridge University Press, 2001.
- [24] N. Schmitt, D. Schmitt, C. Clapham, Developing and Exploring the Behaviour of Two New Versions of the Vocabulary Levels Test, *Language Testing*, Vol. 18, No. 1, 2001, pp. 55-88.
- [25] S. Webb, Y. Sasao, O. Ballance, The Updated Vocabulary Levels Test: Developing and Validating Two New Forms of the VLT, *ITL-International Journal of Applied Linguistics*, Vol. 168, No. 1, 2017, pp. 33-69.
- [26] J. Read, Measuring the Vocabulary Knowledge of Second Language Learners, *RELC Journal*, Vol. 19, No. 2, 1988, pp. 12-25.

- [27] T. Kamimoto, Local Item Dependence on the Vocabulary Levels Test Revisited, *Vocabulary Learning and Instruction*, Vol. 3, No. 2, 2014, pp. 56-68.
- [28] B. Culligan, A Comparison of Three Test Formats to Assess Word Difficulty, *Language Testing*, Vol. 32, No. 4, 2015, pp. 503-520.
- [29] N. Schmitt, P. Nation, B. Kremmel, Moving the Field of Vocabulary Assessment Forward: The Need for More Rigorous Test Development and Validation, *Language Teaching*, Vol. 53, No. 1, 2020, pp. 109-120.
- [30] I. S. P. Nation, D. A. Beglar, Vocabulary Size Test, *The Language Teacher*, Vol. 31, No. 7, 2007, pp. 9-13.
- [31] H. Karami, The Development and Validation of a Bilingual Version of the Vocabulary Size Test, *RELC Journal*, Vol. 43, No. 1, 2012, pp. 53-67.
- [32] D. Beglar, A Rasch-based Validation of the Vocabulary Size Test, *Language testing*, Vol. 27, No. 1, 2010, pp. 101-118.
- [33] B. Laufer, C. Elder, K. Hill, P. Congdon, Size and Strength: Do We Need Both to Measure Vocabulary Knowledge?, *Language Testing*, Vol. 21, No. 2, 2004, pp. 202-226.
- [34] B. Laufer, Z. Goldstein, Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness, *Language Learning*, Vol. 54, No. 3, 2004, pp. 399-436.
- [35] T. A. Levitzky, B. Laufer, Z. Goldstein, The New Computer Adaptive Test of Size and Strength (CATS): Development and Validation. *Language Assessment Quarterly*, Vol. 16, No. 3, 2019, pp. 345-368.
- [36] C. Browne, The New General Service List: Celebrating 60 years of Vocabulary Learning, *The Language Teacher*, Vol. 37, No. 4, 2013, pp. 13-16.
- [37] T. Stoeckel, T. Ishii, P. Bennett, A Japanese-English Bilingual Version of the New General Service List Test, *JALT Journal*, Vol. 40, No. 1, 2018, pp. 5-21.
- [38] T. Stoeckel, S. McLean, P. Nation, Limitations of Size and Levels Tests of Written Receptive Vocabulary Knowledge, *Studies in Second Language Acquisition*, Vol. 43, No. 1, 2021, pp. 181-203.