

NHẬN THỨC VÀ THỰC TRẠNG VỀ KIỂM TRA, ĐÁNH GIÁ TRONG DẠY HỌC CỦA GIÁO VIÊN TIẾNG ANH

Nguyễn Việt Hùng

Trường Quốc tế - Đại học Quốc gia Hà Nội

Email: viethungvnu@vnu.edu.vn

Article history

Received: 25/4/2024

Accepted: 14/5/2024

Published: 20/7/2024

Keywords

Awareness, testing, assessment, teachers, primary school, middle school, secondary school

ABSTRACT

Assessment in teaching in general and English teaching in particular can greatly affect teaching results. Therefore, the National Foreign Language Project 2020 has considered improving the foreign language testing and assessment competence for teachers as one of the pivotal tasks of the Project in the expansion duration from 2017- 2025. By means of a cross-sectional survey carried out in the summer of 2019, systematic random sampling with a sample size of 3103 Vietnamese English teachers at all 3 levels (primary school, middle school, secondary school) by regions (6 key areas in the country, including (1) Northern Mountain Region, (2) Northern Delta Region, (3) Central Region, (4) Central Highlands Region, (5) Southeast Region, and (6) Southwest Region), using quantitative statistical analysis tools, this article is to find out the perception of English teachers regarding their knowledge of testing and assessment, and the actual implementation of assessment in their daily teaching.

1. Mở đầu

Về mặt lí luận giáo dục ngoại ngữ, việc kiểm tra, đánh giá (KT-ĐG) đóng một vai trò quan trọng, và có một tác động rất lớn đối với quá trình dạy học, vì nó cung cấp thông tin để cải thiện việc dạy học liên tục (Lambert & Lines, 2000; Pierce, 2002; Brown, 2004). Việc nhận thức đúng về vai trò của KT-ĐG và hiểu rõ các sử dụng các hoạt động KT-ĐG khác nhau trong lớp học đối với GV tiếng Anh là thực sự cần thiết. Vì nó không chỉ giúp cho GV đánh giá đúng người học mà còn có tác dụng xây dựng động cơ học tập cho người học (Pennington & Hoekje, 2014; Hoque, 2016). Từ những phát hiện đó, các nhà quản lí giáo dục có thể biết được GV tiếng Anh cần các khóa tập huấn nào để nâng cao chất lượng dạy học, từ đó có kế hoạch cụ thể xác thực cho việc điều chỉnh chính sách về sử dụng nhân lực cũng như bồi dưỡng nguồn nhân lực cho phù hợp mà còn giúp cho các bên liên quan, đặc biệt là GV có nhận thức đúng về những điểm mạnh, yếu của bản thân về năng lực khảo thí để chủ động tự bồi dưỡng và thay đổi trong cách dạy học của mình (Pennington & Hoekje, 2014; Kumaravadelu, 2006; Johnson & Golombek, 2011). Từ ý nghĩa đó, nghiên cứu này nhằm mục đích trả lời 3 câu hỏi sau: (1) Ý kiến đánh giá của GV về vai trò của KT-ĐG đối với hoạt động dạy học tiếng Anh? (2) Khả năng và mức độ sử dụng thường xuyên một số hoạt động KT-ĐG của GV tiếng Anh? (3) Ý kiến đánh giá giữa các nhóm GV về vai trò của KT-ĐG có khác nhau không?

2. Kết quả nghiên cứu

2.1. Một số vấn đề lí luận về kiểm tra, đánh giá trong dạy học ngoại ngữ

Nghiên cứu nhận thức của GV về các yếu tố tác động tới quá trình dạy học từ lâu đã được quan tâm rộng rãi trên thế giới. Năng lực và kiến thức về KT-ĐG của GV tiếng Anh ngày càng được quan tâm, và được xác định là yếu tố quan trọng ảnh hưởng tới kết quả dạy học của GV; vì lí do đó mà Đề án Ngoại ngữ quốc gia 2020 (NNQG2020) đã xác định việc bồi dưỡng kiến thức KT-ĐG là một nhiệm vụ quan trọng trong giai đoạn điều chỉnh Đề án giai đoạn 2017-2025 theo Quyết định 2080/QĐ-TTg của Thủ tướng Chính phủ (2008, 2017). Vì vậy, nhận thức của GV tiếng Anh về KT-ĐG là một nhân tố khám phá nên được quan tâm và phân tích, để từ đó tìm ra GV tiếng Anh đang thực sự nhìn nhận vai trò của KT-ĐG tiếng Anh là gì, thực trạng GV đang áp dụng thực tế trên lớp như thế nào.

Theo Pierce (2002), đánh giá là một phần quan trọng của bất kì hoạt động học tập và giảng dạy nào, nó không chỉ giúp đưa ra các quyết định giảng dạy hằng ngày và giúp chẩn đoán điểm mạnh và điểm yếu của HS liên quan đến việc giảng dạy trong lớp, và cung cấp phản hồi cụ thể cho HS để hỗ trợ việc học của họ. Đánh giá cũng cung cấp thông tin phản hồi ngay lập tức cho GV để định hình thực hành giảng dạy phù hợp với phong cách học tập của HS (Graves, 2009; Gupta, 1999). Để đánh giá thành tích của HS và xác định điểm, GV nên sử dụng các bài kiểm tra

khác nhau. Các bài kiểm tra, bài thi và mô hình đánh giá là những công cụ quan trọng, được sử dụng như một thước đo cho quá trình học tập (Richards, 2010; Smith, 1991; Wall, 2005).

Việc KT-ĐG luôn có tác động ngược trở lại (washback effect) đối với việc dạy và học, từ đó tác động tới kết quả chung của khóa học (Alderson & Wall, 1993; Barnes, 2016; Taylor, 2005). Việc tác động có thể theo chiều tích cực hoặc tiêu cực (Hoque, 2011, 2016; Pierce, 2002). Nhiều khi GV thay đổi phương pháp giảng dạy và cách tiếp cận mới để phù hợp với yêu cầu của bài kiểm tra, họ dạy HS của bất cứ thứ gì họ thích dạy, lựa chọn các chương và chủ đề được ưu tiên trong việc giảng dạy và học tập trên lớp, họ thậm chí bỏ qua một số bài của sách giáo khoa và ít ra đề trong bài kiểm tra (Hoque, 2011). Trong nghiên cứu của Watanabe (2004), bỏ qua chương trình giảng dạy, GV đã sử dụng các loại tài liệu tự làm khác nhau để phù hợp với bài kiểm tra mục tiêu. Nhiều GV phụ thuộc nhiều vào các bài kiểm tra có sẵn trong các sách tham khảo bán thương mại (Hayes & Read, 2004; Hoque, 2016).

2.2. Phương pháp nghiên cứu

Nghiên cứu được thực hiện dựa trên phương pháp nghiên cứu khảo sát thực tiễn, sử dụng bảng câu hỏi khảo sát trực tuyến (tại đường link: <http://ulis.vnu.edu.vn/phieu-hoi-giao-vien-ngoai-ngu-pho-thong-ve-hoat-dong-day>), đảm bảo các quy trình chuẩn hóa từ việc lựa chọn bảng hỏi, chuẩn hóa bảng hỏi, chọn mẫu, thu thập số liệu, và phân tích số liệu. Bảng câu hỏi có 12 câu, được thiết kế thành nhiều câu hỏi nhỏ (item) theo thang Likert 5 bậc nhằm phân tích sâu về nhận thức của GV tiếng Anh về kiến thức KT-ĐG, và thực tế thực hiện việc KT-ĐG trong việc dạy học hằng ngày. Bảng câu hỏi được điều tra thử nghiệm, kết quả phân tích độ tin cậy và độ hiệu lực đều rất tốt. Về độ tin cậy Cronbach's Alpha (α), tất cả các câu hỏi gồm nhiều item nhỏ theo thang 5 bậc của bảng hỏi đều đạt giá trị lớn hơn 0.6. Về độ hiệu lực, hầu hết các câu hỏi gồm nhiều item nhỏ theo thang 5 bậc của bảng hỏi đều đồng cấu trúc vì phép kiểm định KMO and Bartlett's Test đều cho kết quả sig < 0.05 và mỗi miền đều triết xuất được 1 factor, không có miền đo hay bảng hỏi nào triết xuất ra 2 factor trở lên.

Số liệu điều tra được xử lý bằng phần mềm SPSS for Windows phiên bản 22.0. Các thông số được sử dụng trong nghiên cứu này bao gồm: (1) các chỉ số thống kê mô tả như tần số tuyệt đối (các số đếm) và tần số tương đối (số phần trăm) đối với các biến dạng số và biến dạng chuỗi với các nhóm hạng không thứ bậc, điểm trung bình cộng, điểm trung vị (median), độ lệch chuẩn (SD); và (2) các chỉ số thống kê suy luận như so sánh giá trị trung bình (compare means) giữa 2 nhóm (T-Test), so sánh giá trị trung bình với 3 nhóm trở nên, phép phân tích phương sai một yếu tố (ANOVA). Đối với các câu có nhiều item nhỏ theo thang 5 bậc cần kiểm định (như Q2, Q7, Q10...), nếu phân phối giá trị tần suất và % không chuẩn hoặc kết quả kiểm định đồng nhất phương sai trong ANOVA không cho giá trị sig > 0.05 thì sử dụng phép kiểm định Kruskal Wallis Test. Đối với các phép so sánh giá trị trung bình của 2 nhóm, phép kiểm định t về độc lập giữa hai mẫu (Independent Samples T-Test) cho biết đối với một nhóm đơn thì trung bình của một nhóm chủ thể này có khác với trung bình của các nhóm chủ thể khác không. Các giá trị trung bình được coi là có ý nghĩa về mặt thống kê khi t-test của phân tích biến thiên có giá trị vượt ngưỡng thống kê với xác suất sig < 0.05. Đối với các biến cần kiểm định (như Q2, Q7, Q10), nếu phân phối giá trị tần suất và % không chuẩn thì sử dụng phép kiểm định Mann Whitney Test.

2.3. Quan niệm của giáo viên tiếng Anh về vai trò của kiểm tra, đánh giá

Ý kiến đánh giá của GV về vai trò của KT-ĐG đối với hoạt động dạy học ngoại ngữ được thể hiện qua bảng 1:

Bảng 1. Tỷ lệ các mức độ đồng ý đối với câu hỏi Q1

	Không đồng ý (1.0-2.9)		Đồng ý (3.0-3.9)		Hoàn toàn đồng ý (4.0-5.0)	
	Tần suất	Phần trăm (%)	Tần suất	Phần trăm (%)	Tần suất	Phần trăm (%)
Q1	69	2.22%	1178	37.96%	1856	59.81%

Theo bảng 1, một nét nổi bật dễ nhận thấy là GV đánh giá cao vai trò của KT-ĐG trong dạy học ngoại ngữ. Theo đó, có tới gần 100% GV tham gia khảo sát có cùng quan điểm như. Có thể tính ra con số trung bình chung về tỷ lệ các mức độ đồng ý đối với các nội dung về vai trò của KT-ĐG là rất cao 97,77% (bao gồm 37,96% đồng ý, và 59,81% hoàn toàn đồng ý), chỉ có 2,22%.

Để phân tích sâu hơn về đánh giá của GV về vai trò của KT-ĐG, các số liệu thống kê mô tả đối với từng câu hỏi nhỏ (item) được sử dụng gồm giá trị trung bình (TB), độ lệch chuẩn (SD), trung tuyến (Median), mode, độ tin cậy (α) và tương quan (r) giữa các item với câu hỏi chính. Các số liệu mô tả về câu hỏi 1 (Q1) được thể hiện qua bảng 2:

Bảng 2. Số liệu thống kê đối với các item trong câu hỏi 1 (Q1)

Câu	N	TB	SD	Median	Mode	α	r
Q1_1	3103	4.15	0.013	4	4	0.526	0.615**
Q1_2	3103	4.06	0.013	4	4	0.631	0.703**

Q1_3	3103	3.94	0.014	4	4	0.606	0.689**
Q1_4	3103	3.9	0.014	4	4	0.581	0.666**
Q1_5	3103	4.08	0.012	4	4	0.593	0.666**
Q1_7	3103	4.09	0.012	4	4	0.597	0.666**
Q1_8	3103	4.12	0.013	4	4	0.582	0.658**
Q1_9	3103	4.06	0.013	4	4	0.457	0.550**
Q1_10	3103	4.22	0.012	4	4	0.614	0.680**
Q1_11	3103	4.11	0.011	4	4	0.52	0.595**
Q1_12	3103	4.15	0.013	4	4	0.56	0.641**
Q1_13	3103	3.97	0.012	4	4	0.557	0.634**
Q1_15	3103	3.49	0.018	4	4	0.257	0.408**
Q1	3003						1

(Ghi chú: min= 1, max= 5)

Theo bảng 2, các tiêu chí Q1.1 và Q1.12 (cả hai đều có điểm trung bình giống nhau, TB = 4.15) được GV đánh giá cao nhất, nghĩa là họ cho rằng “Đào tạo ngoại ngữ hiệu quả là tích hợp được ba thành tố cơ bản và quan trọng nhất là giảng dạy, học tập, và KT-ĐG” (Q1.1), và “Hiệu quả đào tạo tiếng Anh tại Việt Nam sẽ cao hơn khi việc KT-ĐG được kết hợp chặt chẽ với khâu xây dựng chương trình và lựa chọn giáo trình” (Q1.12). Tuy nhiên nhận thức về vai trò của KT-ĐG trong câu hỏi Q1.10 (Việc KT-ĐG hoàn toàn phù hợp với nội dung giáo trình tài liệu giảng dạy cho lớp học, cấp học) lại có vẻ đi ngược lại đánh giá ở câu Q1.12 khi GV đánh giá với điểm số rất cao, TB = 4.22, cho nhận định rằng việc KT-ĐG hoàn toàn phù hợp với nội dung giáo trình tài liệu giảng dạy. Kết quả trên cho thấy hầu hết GV khẳng định rằng KT-ĐG là 1 trong những thành tố quan trọng nhất trong dạy học ngoại ngữ, và hiện tại việc KT-ĐG ở các cấp học chưa hoàn toàn phù hợp với nội dung chương trình giáo trình, đặc biệt là bài thi tốt nghiệp THPT Quốc gia.

Như vậy, chúng tôi nghiêng về nhận định của câu hỏi Q1.12 hơn là nhận định của câu hỏi Q1.10. Tiếp theo sau các nhận định với điểm số cao của ba câu hỏi trên, chúng ta thấy các câu hỏi Q1.8 (Việc KT-ĐG hoàn toàn phù hợp chương trình khung môn ngoại ngữ và chuẩn đầu ra do Bộ GD-ĐT ban hành cho từng lớp học, cấp học), với TB = 4,12, và tiếp ngay sau đó là Q1.11 (Việc KT-ĐG hoàn toàn phù hợp với các hoạt động dạy học trên lớp), với TB = 4,11, cũng có những điểm số rất cao.

Các item có điểm số đánh giá thuộc nhóm thấp hơn (TB < 4) là Q1.3 (KT-ĐG là công việc bắt buộc phải hoàn thành như các công việc khác trong chương trình đào tạo), Q1.4 (KT-ĐG là công cụ đo kết quả học tập của người học theo chuẩn đầu ra đã được xác định), và Q1.15, trong đó nội dung “KT-ĐG để cấp bằng tốt nghiệp hết khóa học, cấp học” (Q1.15), với TB = 3,49. Qua các điểm số này có thể nhận thấy, GV không cho rằng vai trò của KT-ĐG chỉ dừng lại ở việc xét tốt nghiệp. Nghĩa là họ đều nhận thấy tầm quan trọng của KT-ĐG trong việc nâng cao chất lượng dạy học và điều chỉnh quá trình dạy học ngoại ngữ.

2.4. Khả năng và mức độ sử dụng thường xuyên một số hoạt động kiểm tra, đánh giá của giáo viên tiếng Anh

Ý kiến đánh giá của GV về mức độ sử dụng thường xuyên một số hoạt động KT-ĐG (Câu hỏi 8) được thể hiện qua bảng 3.

Bảng 3. Tần suất tự đánh giá về khả năng thực hiện những hoạt động KT-ĐG

Các lựa chọn		8.1	8.2	8.3	8.4	8.5	8.6
Giá trị	Không bao giờ	45	358	19	7	102	458
	Ít khi	114	853	157	53	410	770
	Thỉnh thoảng	492	1080	791	378	1075	953
	Thường xuyên	1551	687	1731	1815	1213	797
	Rất thường xuyên	901	125	405	850	303	125
	Tổng cộng	3103	3103	3103	3103	3103	3103

Theo bảng 3, đa số GV đều cho rằng mình có thực hiện, thường xuyên và rất thường xuyên sử dụng một số hoạt động KT-ĐG trong dạy học ngoại ngữ, như hoạt động 8.1 (Tự chuẩn bị bài kiểm tra trên lớp), hoạt động 8.3 (Phân tích kết quả các bài thi kiểm tra của HS), hoạt động 8.4 (Nhận xét, phản hồi lại cho HS về bài thi, kiểm tra), hoạt động 8.5 (Sử dụng cách kiểm tra chéo giữa các HS/sinh viên hoặc giữa các GV), hoạt động 8.6 (Sử dụng khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam trong KT-ĐG). Tiêu chí có mức độ đồng ý thấp nhất là 8.2 (Sử dụng bài kiểm tra có sẵn trong sách). Kết quả này thể hiện rằng GV đọc khá kỹ câu hỏi 8 và trả lời các tiêu chí với sự cẩn thận nhất định.

Tuy nhiên mức độ thực hiện thường xuyên hoạt động nào đó có thể không hoàn toàn tương quan thuận với khả năng thực hiện thành thạo hoạt động trên (Câu hỏi 9). Các tiêu chí nổi bật với hai lựa chọn này thuộc về Q9.5, Q9.3, Q9.6, và Q9.2, nghĩa là GV cho rằng họ có khả năng “Sử dụng cách kiểm tra chéo giữa các HS/sinh viên hoặc giữa các GV” (Q9.5), “Phân tích kết quả các bài thi kiểm tra của HS” (Q9.3), “Sử dụng khuônng năng lực ngoại ngữ 6 bậc dùng cho Việt Nam trong KT-ĐG” (Q9.6), và “Sử dụng bài kiểm tra có sẵn trong sách” (Q9.2).

Khả năng thực hiện các kỹ năng trong một hoạt động KT-ĐG theo một chu trình khép kín được GV tự đánh giá trong Câu hỏi 7 của bảng hỏi. Bảng 4 sau đây sẽ mô tả trực quan kết quả này:

Bảng 4. Tỷ lệ các mức độ đồng ý đối với câu hỏi Q7

	Không thể thực hiện được (1.0-2.9)		Có thể thực hiện được (3.0-3.9)		Có thể thực hiện thành thạo (4.0-5.0)	
	Tần suất	Phần trăm (%)	Tần suất	Phần trăm (%)	Tần suất	Phần trăm (%)
Q7	734	23.65%	1748	56.33%	621	20.01%

Theo bảng 4, một nét nổi bật dễ nhận thấy là GV đều khá tự tin với khả năng thực hiện các hoạt động KT-ĐG. Cụ thể đại diện cho nhóm ý kiến “làm được” với lựa chọn thuộc về item Q7.2, Q7.3, Q7.4, Q7.5, Q7.6, Q7.7, Q7.8, Q7.9, Q7.24, Q7.25 và Q7.26 chiếm tỉ lệ cao nhất so với các nhóm ý kiến khác đánh giá về từng item trong câu hỏi 7, trên 50%, nhóm ý kiến “làm khá thành thạo” với lựa chọn thuộc về item Q7.1, Q7.2, Q7.7, Q7.9, Q7.12, Q7.13, Q7.14, Q7.15, Q7.22 chiếm tỉ lệ cao thứ hai. Có thể tính ra con số trung bình chung về tỉ lệ đồng ý đối với Khả năng sử dụng các kỹ năng KT-ĐG là gần 80% (bao gồm 56,33% chọn “Có thể thực hiện được”, và 20,01% chọn “Có thể thực hiện thành thạo”), có 23,65% chọn “Không thể thực hiện được”.

Các tiêu chí nổi bật thuộc về Q7.12 (Phát đề thi kiểm tra cho HS/sinh viên để đảm bảo tính chính xác và khách quan), Q7.13 (Hướng dẫn HS/sinh viên các nội quy quy chế trong khi làm bài thi kiểm tra), Q7.14 (Thu bài thi kiểm tra của HS/sinh viên đảm bảo tính chính xác), Q7.15 (Niêm phong, bảo mật đề thi, bài kiểm tra của HS/sinh viên), và Q7.22 (Chấm điểm cho các bài thi kiểm tra của HS/sinh viên dưới dạng giấy).

Đứng sau nhóm top đầu là nhóm top giữa với các kỹ năng được GV tự nhận là có thể thực hiện gồm các tiêu chí Q7.1 (Thiết kế các bài kiểm tra để đo khả năng, kiến thức, kỹ năng của HS), Q7.2 (Xác định các tiêu chí cần đánh giá trước khi thiết kế bài thi kiểm tra), Q7.7 (Kiểm tra độ khó của bài kiểm tra xem có phù hợp để sử dụng cho HS lớp thầy/cô hay không), Q7.9 (Chỉnh sửa bài kiểm tra sau khi kiểm tra thử cho phù hợp), Q7.17 (Thảo luận, thống nhất về tiêu chí đánh giá, thang điểm, trọng số cho từng bài thi kiểm tra), Q7.20 (Thảo luận nhóm chấm về điểm các bài thi kiểm tra).

Nhóm các tiêu chí đứng ở top cuối với các kỹ năng được GV tự nhận là có thể thực hiện ít thành thạo nhất gồm các tiêu chí Q7.3 (Kiểm tra thử để loại bỏ hoặc điều chỉnh những câu hỏi không phù hợp), Q7.4 (Kiểm tra tính ổn định của kết quả bài kiểm tra đối với những đối tượng HS giống nhau), Q7.8 (Đánh giá tính chuẩn xác của các bài thi do người khác thiết kế), Q7.10 (Nhận diện, kiểm tra thí sinh trước và trong phòng thi), Q7.16 (Đánh số, cắt phách các bài thi kiểm tra), Q7.18 (Tập huấn cách cho điểm bài thi nói của HS/sinh viên), Q7.25 (Dùng số liệu thống kê mô tả để xem xét chất lượng các bài kiểm tra).

Để phân tích sâu hơn về khả năng thực hiện các kỹ năng trong 1 hoạt động KT-ĐG khép kín của GV, các số liệu thống kê mô tả đối với từng câu hỏi nhỏ (item) được sử dụng gồm giá trị TB, độ lệch chuẩn (SD), trung tuyến (Median), mode, độ tin cậy (α) và tương quan (r) giữa các item với câu hỏi chính. Các số liệu mô tả về câu hỏi 7 (Q7) được thể hiện qua bảng 5 dưới đây:

Bảng 5. Số liệu thống kê đối với các item trong câu hỏi 7 (Q7)

Câu	N	TB	SD	Median	Mode	α	r
Q7_1	3103	3.46	0.777	3	3	0.677	0.702**
Q7_2	3103	3.42	0.737	3	3	0.717	0.738**
Q7_3	3103	3.16	0.789	3	3	0.637	0.665**
Q7_4	3103	3.12	0.792	3	3	0.67	0.695**
Q7_5	3103	3.26	0.759	3	3	0.722	0.744**
Q7_6	3103	3.28	0.753	3	3	0.747	0.767**
Q7_7	3103	3.33	0.749	3	3	0.75	0.769**
Q7_8	3103	3.18	0.778	3	3	0.713	0.735**
Q7_9	3103	3.33	0.793	3	3	0.744	0.764**
Q7_10	3103	3.28	0.892	3	3	0.735	0.759**
Q7_11	3103	3.47	0.884	3	3	0.778	0.798**
Q7_12	3103	3.61	0.879	3	3	0.798	0.817**
Q7_13	3103	3.67	0.863	4	3	0.799	0.818**
Q7_14	3103	3.7	0.88	4	3	0.792	0.811**

Q7_15	3103	3.59	0.931	3	3	0.768	0.791**
Q7_16	3103	3.28	1.002	3	3	0.67	0.703**
Q7_17	3103	3.43	0.831	3	3	0.813	0.829**
Q7_18	3103	3.1	0.909	3	3	0.651	0.682**
Q7_19	3103	3.33	0.846	3	3	0.792	0.810**
Q7_20	3103	3.32	0.852	3	3	0.79	0.809**
Q7_21	3103	3.34	0.98	3	3	0.71	0.739**
Q7_22	3103	3.67	0.871	4	3	0.77	0.791**
Q7_23	3103	3.24	0.909	3	3	0.655	0.686**
Q7_24	3103	3.3	0.829	3	3	0.784	0.803**
Q7_25	3103	3.2	0.864	3	3	0.73	0.754**
Q7_26	3103	3.28	0.831	3	3	0.737	0.759**
Q7	3103	3.36	0.64	3.19	3	0.97	1

Qua bảng 5, ta thấy GV tự nhận điểm số đánh giá ở mức trung bình về khả năng thực hiện các kỹ năng trong 1 hoạt động KT-ĐG khép kín, hầu hết điểm trung bình TB nằm trong khoảng 3.1 - 3.4, điểm trung bình của Câu 7 là TB = 3.36, hầu hết trung tuyến của các item và tất cả mode của các item đều là 3. Phép phân tích tương quan nhị biến cho thấy tất cả các item có tương quan thuận và khá chặt với Câu hỏi 7 (r chủ yếu nằm trong từ 0.7 - 0.8). Chỉ số độ tin cậy của từng item khá cao (α nằm trong khoảng từ 0.6 - 0.8), không có tiêu chí nào bị loại bỏ vì độ tin cậy thấp.

Tuy nhiên có một điểm nổi bật không mong đợi từ kết quả trên là những kỹ năng được các chuyên gia phương pháp mong đợi GV cần thực hiện tốt như khả năng “Kiểm tra thử để loại bỏ hoặc điều chỉnh những câu hỏi không phù hợp” (Q7.3), khả năng “Kiểm tra tính ổn định của kết quả bài kiểm tra” (Q7.4), khả năng “Đánh giá tính chuẩn xác của các bài thi do người khác thiết kế” (Q7.8), khả năng “Dùng số liệu thống kê mô tả để xem xét chất lượng các bài kiểm tra” (Q7.25.) và các khả năng đo độ phân biệt, độ khó, độ giá trị của bài kiểm tra lại nằm ở tốp cuối. Các khả năng quan trọng khác như “Xác định các tiêu chí cần đánh giá trước khi thiết kế bài thi kiểm tra” (Q7.2), khả năng “đo độ khó của bài kiểm tra xem có phù hợp với HS không” (Q7.7), khả năng “Chỉnh sửa bài kiểm tra sau khi kiểm tra thử cho phù hợp” (Q7.9) lại nằm ở tốp giữa. Những khả năng mà GV cho rằng họ thành thạo nhất lại là những kỹ năng rất đơn giản như Q7.12 (Phát đề thi kiểm tra cho HS), Q7.13 (Hướng dẫn nội quy quy chế làm bài thi kiểm tra), Q7.14 (Thu bài thi kiểm tra), Q7.15 (Niêm phong, bảo mật đề thi), và Q7.22 (Chấm điểm cho các bài thi kiểm tra của HS/sinh viên dưới dạng giấy).

2.5. So sánh đánh giá của giáo viên các cấp về vai trò của kiểm tra, đánh giá

Kết quả đánh giá về vai trò của KT-ĐG giữa 3 nhóm GV (thể hiện trong kết quả trả lời đối với câu hỏi 1 trong bảng hỏi dành cho GV) được thể hiện trực quan trong bảng 6 dưới đây:

Bảng 6. So sánh giá trị trung bình của các nhóm GV về vai trò của KT-ĐG

Đối tượng	N	Mean	SD	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
GV tiểu học	974	3.9972	0.43386	0.01390	3.9699	4.0245	2.54	5.00
GV THCS	1300	4.0173	0.45623	0.01265	3.9925	4.0421	1.38	5.00
GV THPT	829	4.0721	0.48148	0.01672	4.0392	4.1049	1.00	5.00
Total	3103	4.0256	0.45710	0.00821	4.0095	4.0417	1.00	5.00

Theo bảng 6, điểm số đánh giá tầm quan trọng của của KT-ĐG đối với việc dạy học ngoại ngữ tăng dần giữa 3 nhóm GV từ cấp thấp lên cấp cao. GV THPT đánh giá cao nhất về vai trò của KT-ĐG (TB = 4.0721), sau đó là GV THCS (TB = 4.0173), và GV tiểu học là nhóm đánh giá thấp nhất về vai trò của KT-ĐG (TB = 3.9972).

Theo kết quả so sánh ANOVA giữa các nhóm thì đây là những điểm số khác biệt có ý nghĩa vì $\text{Sig.} = 0.002 < 0.05$. Điểm số chênh lệch cụ thể được thể hiện trong bảng 7. Theo bảng 7, chúng ta thấy GV THPT có điểm số đánh giá về vai trò của KT-ĐG cao hơn hai nhóm còn lại, cụ thể cao hơn GV tiểu học là 0.07485 đơn vị trong khoảng tin cậy 95% là $(0.001 \div 0.033)$, cao hơn GV THCS là 0.05477 đơn vị trong khoảng tin cậy 95% là $(0.007 \div 0.015)$. Điều này cho thấy, GV dạy ở cấp học càng cao thì càng đánh giá cao vai trò của KT-ĐG trong dạy học.

Bảng 7. Post Hoc giữa các nhóm

(I) Nam	(J) Nam	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
GV tiểu học	GV THCS	-0.020	0.019	0.299	-0.058	0.018
	GV THPT	-0.07485*	0.022	0.001	-0.117	-0.033
GV THCS	GV tiểu học	0.020	0.019	0.299	-0.018	0.058
	GV THPT	-0.05477*	0.020	0.007	-0.095	-0.015
GV THPT	GV tiểu học	0.07485*	0.022	0.001	0.033	0.117
	GV THCS	0.05477*	0.020	0.007	0.015	0.095

3. Kết luận

Kết quả khảo sát cho thấy, hầu hết GV tiếng Anh khẳng định rằng KT-ĐG là một trong những thành tố quan trọng nhất trong dạy học ngoại ngữ. Qua các điểm số đánh giá có thể nhận thấy các GV không cho rằng vai trò của KT-ĐG chỉ dừng lại ở việc xét tốt nghiệp. Nghĩa là họ đều nhận thấy tầm quan trọng của KT-ĐG trong việc nâng cao chất lượng dạy học và điều chỉnh quá trình dạy học ngoại ngữ. Tuy nhiên, việc KT-ĐG ở các cấp học chưa hoàn toàn phù hợp với nội dung chương trình giáo trình, đặc biệt là bài thi tốt nghiệp THPT Quốc gia. Kết quả đánh giá về vai trò của KT-ĐG giữa 3 nhóm GV tiểu học, THCS và THPT thì GV THPT đánh giá cao nhất về vai trò của KT-ĐG, sau đó là GV THCS, và GV tiểu học là nhóm đánh giá thấp nhất về vai trò của KT-ĐG; và có sự khác biệt có ý nghĩa về chênh lệch giữa GV tiểu học và THPT.

Kết quả khảo sát cũng cho thấy, GV tiếng Anh thường xuyên sử dụng một số hoạt động KT-ĐG trong khi dạy học, nhưng mức độ thực hiện thường xuyên hoạt động nào đó có thể không hoàn toàn tương quan thuận với khả năng thực hiện thành thạo hoạt động trên. Hầu hết GV tự nhận là biết thiết kế các bài KT-ĐG để sử dụng trên lớp; tuy nhiên, khi hỏi về các loại bài kiểm tra cụ thể thì đa số GV thừa nhận có một số các hoạt động GV ít sử dụng trên lớp, điều này có liên quan với khả năng không thực hiện thành thạo hoạt động đó. Qua phỏng vấn, gần như tất cả GV cũng thừa nhận rằng, họ không thể đo độ khó, độ tin cậy, độ phân biệt, độ xác trị của các bài kiểm tra do họ thiết kế, họ cũng không thể thiết kế và đảm bảo các ma trận kiến thức kỹ năng để đảm bảo ra đề kiểm tra bám sát nội dung chương trình học.

Tài liệu tham khảo

- Alderson, C. J., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129. <https://doi.org/10.1093/aapl/14.2.115>
- Barnes, M. (2016). The Washback of the TOEFL iBT in Vietnam. *Australian Journal of Teacher Education*, 41(7), 158-174. <https://doi.org/10.14221/ajte.2016v41n7.10>
- Brown, D. H. (2004). *Language assessment principles and classroom practices*. Pearson Education.
- Graves, K. (2009). The curriculum of second language teacher education. In Burns, A and Richards, J (Eds), *Second Language Teacher Education*. Cambridge: Cambridge University Press, pp. 115-124.
- Gupta, K. (1999). *A practical guide for need assessment*. John Wiley & Sons, Inc.
- Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*, (pp. 97-111). Lawrence Erlbaum Associates.
- Hoque, M. E. (2011). *Washback of the public examination on teaching and learning English as a foreign language (EFL) at the higher secondary level in Bangladesh, Doctoral dissertation*. Jahangirnagar University.
- Hoque, M. E. (2016). Teaching to the EFL curriculum or teaching to the test: An investigation. *The EDRC Journal of Learning and Teaching*, 1(1), 6-30.
- Johnson, K. E., & Golombek, P. R. (Eds.). (2011). *Research on second language teacher education: A sociocultural perspective on professional development*. New York: Routledge.
- Kumaravadivelu, B. (2006). *Understanding language teaching: From method to postmethod*. Mahwah, NJ.: Lawrence Erlbaum.
- Lambert, D., & Lines, D. (2000). *Understanding Assessment: Purposes, Perceptions, Practices*. Routledge Falmer.
- Pennington, M. C., & Hoekje, B. J. (2014). Framing English language teaching. *System*, 46, 163-175.
- Pierce, L. V. (2002). Performance-Based Assessment: Promoting Achievement for Language Learners. *Center for Applied Linguists (ERIC/CLL News Bulletin)*, 26(1), 1-30.
- Richards, J. C. (2010). Competence and Performance in Language Teaching. *RELC Journal*, 41(2), 101-122. <https://doi.org/10.1177/0033688210372953>
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11. <https://doi.org/10.3102/0013189X020005008>
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155. <https://doi.org/10.1093/eltj/cci030>
- Thủ tướng Chính phủ (2008). *Quyết định số 1400/QĐ-TTg ngày 30/9/2008 về Phê duyệt Đề án “Dạy và học ngoại ngữ trong hệ thống giáo dục quốc dân giai đoạn 2008-2020*.
- Thủ tướng Chính phủ (2017). *Quyết định số 2080/QĐ-TTg ngày 22/12/2017 về việc Phê duyệt điều chỉnh, bổ sung Đề án dạy và học ngoại ngữ trong hệ thống giáo dục quốc dân giai đoạn 2017-2025*.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge University Press.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*, (pp. 19-36). Lawrence Erlbaum Associates.