

DÁNH GIÁ ĐỘ TIN CẬY CỦA NGÂN HÀNG ĐỀ THI CĂN CỨ TRÊN KẾT QUẢ THI KẾT THÚC HỌC PHẦN

**Trần Công Nghiệp^{*}, Mai Việt Anh, Nguyễn Thị Thúy Vân
Trường Đại học Kinh tế & Quản trị kinh doanh – ĐH Thái Nguyên**

TÓM TẮT

Độ tin cậy là một tiêu chí quan trọng trong đánh giá một bài thi. Định kỳ đánh giá độ tin cậy của các phương pháp kiểm tra đánh giá là yêu cầu của kiểm định chương trình đào tạo. Việc đánh giá độ tin cậy của cả bộ ngân hàng đề thi, câu hỏi thi bao gồm nhiều hình thức thi như thi trắc nghiệm, thi tự luận, thi vấn đáp là một bài toán khó. Bài báo đề xuất một phương pháp đánh giá độ tin cậy của bộ ngân hàng câu hỏi thi của trường Đại học Kinh tế và Quản trị kinh doanh, Đại học Thái Nguyên theo phương pháp thống kê.

Từ khóa: *Đánh giá, độ tin cậy, ngân hàng câu hỏi thi, ngân hàng đề thi*

ĐẶT VĂN ĐỀ

Kiểm tra đánh giá là một khâu rất quan trọng trong quá trình dạy học, đó là công cụ để người dạy đánh giá mức độ tiếp thu của người học, điều chỉnh kiến thức cung cấp cho người học phù hợp với mức độ tiếp thu. Đối với một bài thi, bốn tiêu chí quan trọng cần phải đạt được là tính giá trị, độ tin cậy, tính khả thi và độ phân biệt. Trong bộ tiêu chí kiểm định chương trình đào tạo ban hành kèm theo Quyết định số 474/QĐ-ĐHTN ngày 22 tháng 5 năm 2012, có tiêu chí 4.4 qui định “Độ tin cậy và tính giá trị của các phương pháp kiểm tra đánh giá được định kỳ thẩm định”. Độ tin cậy của một bài kiểm tra là tính nhất quán mà một bài kiểm tra luôn đánh giá chính xác một đối tượng tại các thời điểm khác nhau *Harrison*. Độ tin cậy là điều kiện cần của các đề thi – kiểm tra để đánh giá chính xác kết quả học tập của người học. Độ tin cậy có thể tăng lên nhờ bổ sung những câu hỏi đồng nhất (tăng độ dài của đề thi) hay tăng độ phân biệt của các câu hỏi [1].

Cho đến hiện tại, có nhiều công trình nghiên cứu về độ tin cậy của các bài thi trắc nghiệm. Tuy nhiên, chưa có nghiên cứu nào đề cập đến độ tin cậy của các bài thi tự luận, bài thi vấn đáp. Theo yêu cầu của tiêu chí kiểm định chương trình đào tạo thì đơn vị vận hành chương trình đào tạo phải thường xuyên thẩm

định độ tin cậy của phương pháp kiểm tra đánh giá. Tại trường Đại học Kinh tế và Quản trị kinh doanh, đối với hình thức thi tự luận và thi trắc nghiệm khách quan, mỗi lần thi, đề thi được lấy ngẫu nhiên bằng phần mềm máy tính từ ngân hàng câu hỏi thi theo kết cấu đề thi do bộ môn đề xuất. Số lượng đề thi tùy thuộc số lượng thí sinh dự thi đảm bảo mỗi ca thi có hai đề thi. Đối với hình thức thi vấn đáp, mỗi lần thi sử dụng toàn bộ các đề thi trong ngân hàng đề thi vấn đáp. Do vậy rất khó đánh giá được hết toàn bộ ngân hàng câu hỏi thi tự luận do được lựa chọn ngẫu nhiên mà có câu hỏi được sử dụng lặp lại nhiều lần nhưng có câu hỏi chưa được chọn lần nào sau nhiều lần thi. Bằng thống kê chúng tôi nhận thấy, sau 5 lần tổ chức thi đối với các học phần có nhiều thí sinh dự thi thì có khoảng trên 60% số câu hỏi có trong ngân hàng câu hỏi đã được sử dụng. Vì vậy chúng tôi đề xuất sử dụng phương pháp thống kê, phân tích kết quả thi để đánh giá độ tin cậy của ngân hàng câu hỏi thi, ngân hàng đề thi của tất cả các học phần đang được giảng dạy tại trường.

LỰA CHỌN PHƯƠNG PHÁP ĐÁNH GIÁ ĐỘ TIN CẬY CỦA NGÂN HÀNG ĐỀ THI

Phương pháp kiểm tra 2 lần dùng để xác định độ tin cậy của một bài thi trắc nghiệm

Phương pháp này do Henning (1987) và Shohamy (1985) đề xuất, theo phương pháp này một bài kiểm tra được kiểm tra hai lần cho cùng một đối tượng sinh viên trong thời

* Tel 0912 967 494, Email: Trancangnghiep@tueba.edu.vn

gian không quá 2 tuần với điều kiện người học không được ôn tập lại. Độ tin cậy của bài kiểm tra trong trường hợp này được tính đơn giản như sau:

$$r_{tt} = R_{1,2}$$

Trong đó: r_{tt} : Độ tin cậy của bài thi

$R_{1,2}$: Sự tương quan giữa kết quả 2 lần kiểm tra trên cùng một đối tượng

Tuy nhiên phương pháp này chưa đủ khách quan vì không đảm bảo chắc chắn là người học không ôn lại giữa hai lần kiểm tra [5].

Phương pháp phân đôi bài kiểm tra để tính độ tin cậy của bài thi trắc nghiệm

Theo phương pháp này bài kiểm tra được chia thành 2 phần bằng nhau, ví dụ một phần gồm toàn câu chẵn, một phần gồm toàn câu lẻ và tiến hành kiểm tra trên cùng một đối tượng. Sau khi chấm riêng biệt từng phần, người ta so sánh tương quan của kết quả. Điểm số của hai phần càng giống nhau nghĩa là bài thi càng có giá trị. Công thức Spearman được sử dụng để tính hệ số tương quan:

$$R_{tt} = \frac{2R_{AB}}{1 + R_{AB}}$$

trong đó: R_{tt} : là hệ số tin cậy theo phương pháp chia đôi

$R_{A,B}$: là hệ số tương quan giữa hai điểm số của hai phần của bài kiểm tra

Để tính hệ số tương quan giữa hai phần của bài kiểm tra ($R_{A,B}$), người ta có thể dùng công thức:

$$R_{AB} = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$$

Trong đó:

$R_{A,B}$: là hệ số tương quan Spearman

D: sự chênh lệch điểm của hai phần bài kiểm tra
n: số lượng bài kiểm tra

Nếu $R_{A,B}$: (0,8 - 1,0) sự tương quan tốt (độ tin cậy cao); (0,6 - 0,8) sự tương quan trung bình (độ tin cậy trung bình); (0,4 - 0,6) sự tương quan kém (độ tin cậy thấp); (0,2 - 0,4) sự tương quan rất kém (độ tin cậy rất thấp).

Tuy nhiên, cả Bachman (1990) và Henning (1987) đều cho rằng độ tin cậy tính theo phương pháp chia đôi có thể không mang lại kết quả chính xác vì độ tin cậy có thể thay đổi tùy theo bố cục của bài kiểm tra [2].

Công thức 20 của Kuder-Richardson để tính độ tin cậy của bài thi trắc nghiệm

Để khắc phục nhược điểm trên Henning (1987) đã đưa ra một số công thức tính độ tin cậy:

$$R_{tt} = \frac{n}{n-1} \left(\frac{S_f^2 - \sum S_i^2}{S_f^2} \right)$$

Trong đó:

R_{tt} : độ tin cậy

n: số câu hỏi trong bài kiểm tra

S_f^2 : bình phương của độ lệch chuẩn (SD)

$\sum S_i^2$: tổng số sự chênh lệch điểm của tất cả các câu hỏi

Công thức 21 của Kuder-Richardson để tính độ tin cậy của bài thi trắc nghiệm

Tuy nhiên công thức 20 của Kuder-Richardson rất khó tính toán. Trong trường hợp không tính được sự chênh lệch điểm của từng câu hỏi, người ta khuyên nên dùng công thức sau:

$$R_{tt} = 1 - \frac{\bar{x} - (K - \bar{x})}{K(SD)^2}$$

Trong đó:

\bar{x} : điểm trung bình chung (Mean)

$(SD)^2$: bình phương của độ lệch chuẩn (SD)

K: số câu hỏi trong bài kiểm tra

Bachman (1990) cho rằng về mặt cơ bản, mặc dù hai công thức trên được tính khác nhau nhưng chúng đều cho những kết quả như nhau vì sự chênh lệch điểm số bằng với độ lệch chuẩn. Henning (1987) cho rằng độ tin cậy của bài thi nằm trong khoảng từ 0 đến 1 với giá trị càng cao thì bài thi càng có độ tin cậy cao.

KẾT QUẢ PHÂN TÍCH

Để đánh giá được phần nào độ tin cậy của các đề thi kết thúc học phần của Trường Đại học Kinh tế và Quản trị kinh doanh - ĐH Thái Nguyên trong những năm học gần đây, nhóm tác giả tiến hành phân tích, đánh giá kết quả thi kết thúc học phần của hai năm học gần nhất (năm học 2013-2014; 2014-2015) bằng phương pháp thống kê toàn bộ điểm thi kết thúc học phần của tất cả các học phần trong hai năm học đó và so sánh điểm của các học phần tại các thời điểm khác nhau.

Phương pháp phân tích

Dữ liệu được chọn là toàn bộ kết quả thi kết thúc học phần của 2 năm học gồm 2013-2014 và 2014-2015 tổng hợp trên cơ sở các phiếu điểm lưu trữ tại phòng Khảo thí và Đảm bảo chất lượng giáo dục bao gồm 541 lượt học phần với 113.628 lượt thí sinh dự thi. Các học phần có tên giống nhau nhưng được giảng ở lớp thường và lớp chất lượng cao được phân tích riêng để làm cơ sở so sánh.

Dữ liệu được nhập vào phần mềm Excel, được xử lý bằng phần mềm MATLAB. Các chỉ tiêu thống kê mô tả được tính toán bao gồm: Số lượng bài thi, điểm trung bình, độ lệch chuẩn, số bài thi ở 2 cận biên của miền điểm là điểm 0-1 và điểm 9-10. Các mức đánh giá theo hệ thống tín chỉ tương đương bao gồm F (điểm <4), D (điểm từ 4-5,4), C

(điểm từ 5,5- 6,9), B (điểm từ 7 đến 8,4) và A (điểm từ 8,5 trở lên) được tính tỉ lệ phần trăm làm cơ sở để đánh giá mức độ tập trung của miền kết quả của từng học phần.

Nhóm tác giả đã phân tích kết quả thi của các học phần và xem xét đến việc có đảm bảo khoa học phân bố của kết quả điểm tương đối cân đối giữa các mức đánh giá hay không với đề xuất mức điểm trung bình nên ở khoảng 5,5 với độ lệch chuẩn khoảng từ 1,5 đến 1,8. Miền kết quả này tương đương với khoảng trên 60% bài thi có điểm trung bình từ tích D đến tích B, đảm bảo phân loại được sinh viên.

Tổng số 541 lượt học phần, 113.628 bài thi bao gồm cả thi vấn đáp, thi trắc nghiệm và thi tự luận, thực hành. Trong đó học phần có số bài thi cao nhất là Giáo dục thể chất có 1468 bài thi.

Phân tích mức điểm trung bình cao nhất và thấp nhất trong 02 năm học 2013-2014; 2014-2015

Điểm trung bình cao nhất và thấp nhất các học kỳ được thể hiện ở bảng 2.

Số liệu bảng 2 cho thấy điểm trung bình của một số học phần còn chưa được phân bổ chuẩn với những học phần có mức điểm trung bình quá cao: 8,84 điểm, trong khi đó có học phần lại quá thấp 2,83 điểm.

Bảng 1: Báo cáo tổng hợp chung số liệu của 2 năm học

Stt	Năm học	Học kỳ	Lượt môn	Lượt TS dự thi	Hình thức thi (lượt môn)			
					Viết	Vấn đáp	Thực hành	TNMT
1		HK I	96	14836	59	20	5	9
2	2013-2014	HK II	103	30788	69	20	2	12
3		HK Phụ	55	7133	32	14	3	4
4		HK I	133	38537	85	35	2	11
5	2014-2015	HK II	82	18355	60	14	1	6
6		HK Phụ	72	3979	39	24	3	5
Tổng			541	113.628				

Bảng 2: Bảng tổng hợp học phần điểm trung bình cao nhất và thấp nhất

STT	NĂM HỌC	HỌC KỲ	Điểm Trung bình (diểm số)		Ghi chú
			Cao nhất	Thấp nhất	
1	2013-2014	HK I	(8,84)	(3,11)	
2		HK II	(8,38)	(3,46)	
3	2014-2015	HK I	(8,65)	(3,23)	
4		HK II	(8,35)	(2,83)	

Phân tích cụ thể theo từng học kỳ, phân điểm trung bình thành 3 khoảng ta có số liệu cụ thể:

Bảng 3: Kết quả phân tích điểm trung bình theo từng khoảng cụ thể

STT	NĂM HỌC	HỌC KỲ	ĐVT: Lượt môn		
			Điểm TB từ 0 - 4,98	Điểm TB từ 5,0 - 5,59	Điểm từ 5,60-9,0
1		HK I	27	13	56
2	2013-2014	HK II	30	19	54
3		HK Phụ	11	9	35
Tổng			68	37	149
4		HK I	26	27	80
5	2014-2015	HK II	25	14	43
6		HK Phụ	36	7	29
Tổng			70	48	169
					287

Qua bảng số liệu phân tích trên cho thấy số học phần có mức điểm trung bình từ 5,60 đến 9,0 luôn chiếm tỷ lệ rất cao so với các mức trung bình từ 5,0 đến 5,59. Mức điểm trung bình này cũng cho thấy hiện tại số học phần có điểm trung bình quá thấp hoặc quá cao luôn chiếm tỷ lệ lớn, tạo ra sự mất cân bằng giữa các mức điểm trong nhiều học phần.

Phân tích theo số bài thi đạt điểm quá thấp và quá cao

Báo cáo tổng hợp số liệu của các học phần có số bài thi đạt điểm quá thấp (0-1) điểm và quá cao (9-10) điểm.

Bảng 4: Bảng tổng hợp học phần có điểm thấp nhất (0-1) và cao nhất (9-10)

ĐVT: Số bài thi

STT	NĂM HỌC	Học kỳ	SỐ BÀI THI ĐIỂM 0-1 và ĐIỂM 9-10		Ghi chú
			Thấp nhất (diểm 0-1)	Cao nhất (diểm 9-10)	
1	2013-2014	HK I	115/1356 - (8,5%) - Nhũng NLCBCNMLN	141/1468 (9,6%) - GDTC 1	
2		HK II	143/668-(21,5%) - Kinh tế lượng; 130/1364 - (9,5%) - Nhũng NLCBCNMLN 2	256/1363- (18,8%) - Lý thuyết XSTK	
3	2014-2015	HK I	131/1403- (9,3%) GDTC2	510/845 – (60,3%) Kinh tế phát triển	
4		HK II	329/1077 – (30,5%) Tư tưởng HCM	107/1116 – (9,5%) Lý thuyết SXTK	

Nhìn từ bảng trên cho thấy Môn Kinh tế lượng và môn Những Nguyên lý cơ bản của chủ nghĩa Mác Lê-nin HP2 có số bài thi đạt điểm 0-1 rất cao và có tính lặp lại ở các năm. Môn Kinh tế Phát triển có điểm đạt 9-10 với số bài quá nhiều 510/845 (60%).

Phân tích theo tiêu chí điểm đạt tương đương tích F (0 - 3,9) và tích A (8,5-10)

Tổng hợp số liệu chung tại các khoa, số học phần có điểm đạt tương đương tích F và tích A từ 40% trở lên được cụ thể hóa tại bảng sau:

Bảng 5: Bảng tổng hợp số lượt môn có 40% trở lên điểm tương đương tích A, tích F

STT	Khoa	Năm học 2013 - 2014		Năm học 2014 - 2015	
		Số lượt môn 40% có điểm F trở lên	Số lượt môn 40% có điểm A trở lên	Số lượt môn 40% có điểm F trở lên	Số lượt môn 40% có điểm A trở lên
1	Kế toán	1	0	1	1
2	Kinh tế	4	2	4	3
3	Khoa học cơ bản	9	5	8	3
4	MKT-TM-DL	0	2	1	2
5	Ngân hàng Tài chính	4	3	7	0
6	Quản trị Kinh doanh	4	0	3	0
7	Quản lý Luật Kinh tế	0	2	0	2
Tổng		22	14	24	11

Qua bảng trên cho thấy rõ khoa Khoa học Cơ bản luôn có số lượt môn đạt điểm tương đương F cao nhất trong cả 02 năm học. Các khoa gồm khoa Kế toán, Marketing – Thương mại – Du lịch, Quản lý – Luật Kinh tế là những đơn vị có số môn tích F trên 40% thấp nhất so với các khoa khác trong nhà trường. Đối với số lượt môn điểm trên 40% đạt tích A có 02 khoa gồm Kế toán và khoa Quản trị kinh doanh có số lượng rất ít hoặc không có.

Phân tích từ độ lệch chuẩn điểm các học phần

Từ điểm thống kê được, nhóm tác giả tiến hành phân tích độ lệch chuẩn của điểm thi các học phần, kết quả cụ thể như sau:

Bảng 6: Bảng phân tích độ lệch chuẩn

STT	NĂM HỌC	HỌC KỲ	ĐLC từ 0 - 1,49	ĐLC từ 1,5 - 1,8	ĐLC từ 1,81 –trở lên	Tổng
1	2013-2014	HK I	27	21	48	96
2		HK II	33	26	44	103
3		HK Phụ	14	6	35	55
Tổng	2014-2015		74	53	127	254
4		HK I	43	26	64	133
5		HK II	35	19	28	82
6		HK Phụ	26	14	32	72
Tổng			104	62	121	287

Nhìn từ bảng phân tích trên cho thấy, trong cả hai năm học, số học phần có độ lệch chuẩn từ 1,5-1,8 luôn chỉ chiếm tỷ lệ nhỏ hơn nhiều so với 02 nhóm mức độ còn lại. Độ lệch chuẩn từ 1,81 trở lên luôn nhiều gấp 2 lần nhóm học phần có độ lệch chuẩn từ 1,5-1,8. Qua kết quả

ta cũng nhận thấy rằng có sự chênh lệch lớn về giá trị của từng thời điểm đánh giá so với giá trị điểm trung bình.

KẾT LUẬN

Phân tích và đánh giá các câu hỏi thi – kiểm tra là một việc rất cần thiết nhưng còn ít được quan tâm ở nhiều nơi. Hậu quả là các kỳ thi sau vẫn thường lặp lại những sai lầm của các kỳ thi trước. Việc phân tích, đánh giá các câu hỏi sẽ giúp chúng ta hiểu rõ hơn chất lượng của đề thi và chất lượng của học sinh. Tuy nhiên, mức độ phức tạp của việc phân tích, đánh giá các câu hỏi phụ thuộc nhiều vào dạng thức câu hỏi được sử dụng trong kỳ thi.

Thực tế cho thấy rất khó phân tích và đánh giá các câu hỏi tự luận. Rất khó đánh giá độ khó của các câu hỏi này, khó xác định những câu tự luận mơ hồ để loại bỏ hoặc sửa chữa chúng. Trong khi đó, với các câu hỏi trắc nghiệm đã có khá nhiều phần mềm để phân tích đánh giá. Nhờ đó, các câu hỏi trắc nghiệm thường xuyên được nâng cao chất lượng, loại bỏ hoặc sửa chữa được những câu thi rõ ràng, nâng được độ tin cậy và tính giá trị của các đề thi – kiểm tra.

Do không chú ý đến việc phân tích đề thi, không có thử nghiệm trước khi thi nên một số đề thi – kiểm tra quá dễ hoặc quá khó so với năng lực của học sinh. Đề quá dễ dẫn đến kết quả làm bài của học sinh quá tốt nên bị xã hội phê phán là chạy theo thành tích, không phản ánh đúng thực lực của học sinh. Đề quá khó dẫn đến kết quả làm bài của học sinh quá kém nên bị xã hội phê phán là chất lượng giáo dục thấp. Tuy nhiên, đánh giá chất lượng giáo dục mà chỉ nhìn vào điểm số là cách đánh giá rất thiếu chính xác vì điểm số phụ thuộc rất nhiều vào độ khó của đề thi và tính chủ quan của người chấm. Không thể so sánh kết quả hai kỳ thi khác nhau khi các đề thi đó có độ khó khác nhau. Thậm chí, cùng một đề thi nhưng

do những người chấm khác nhau cũng dẫn đến kết quả khác nhau.

Đánh giá chung về độ tin cậy của ngân hàng câu hỏi thi

Qua kết quả phân tích các mức điểm đặc biệt, mức điểm trung bình và độ lệch chuẩn so với điểm trung bình các học phần trong 02 năm học cho thấy có nhiều học phần có tỷ lệ điểm thi quá thấp hoặc quá cao, độ lệch chuẩn so với điểm trung bình còn có khoảng cách lớn.

Kiến nghị

Để nâng cao độ chính xác của phương pháp này, cần phải tiến hành phân tích kết quả thi sau mỗi kỳ thi và nhiều lần phân tích so sánh sẽ cho kết quả độ tin cậy của ngân hàng câu hỏi thi, ngân hàng đề thi chính xác hơn.

Các bộ môn trong Nhà trường phải nghiêm túc triển khai các hoạt động rà soát định kỳ và đánh giá ngân hàng câu hỏi đề thi, để từ đó rút ra những điểm mạnh điểm yếu, những vấn đề cần khắc phục trong khâu xây dựng ngân hàng câu hỏi, ra đề thi.

TÀI LIỆU THAM KHẢO

1. Nguyễn Công Khanh, 2004, *Đánh giá và do lường* trong KHXH, NXB chính trị QG, HN
2. Thorndike & Hagen, *Đo lường và đánh giá* trong tâm lý và giáo dục, Bản dịch của DHQGN
3. Bachman, L.F (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
4. Harrison, A. (1983). *A Language Testing Handbook*. London: Mcmillian Press.
5. Heaton, J.B. (1988). *Writing English Language Tests*. London: Long man
6. Henning, G. (1987). *A Guide to Language Testing*. Cambridge: Newbury House Publishers
7. Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
8. Lado, R. (1961). *Language Testing*. London: Longman.
9. Shohamy, E. (1985). *A Practical Hanbook in Language Testing for the Second Language Teachers*. Tel-Aviv: Tel-Aviv University Press.

SUMMARY**EVALUATION REALITY OF QUESTION BANK
ON BASES MODULES TEST BANK**

Tran Cong Nghiеп^{*}, Mai Viet Anh, Nguyen Thi Thuy Van

College of Economics and Business Administration – TNU

Reality of a test is one of the most important elements of a test evaluation. In educational auditing, it is required to evaluate the reality of testing methods periodically. However, to evaluate the reality of a set of question banks including oral examination, essay examination as well as computer test is difficult. This paper propose the statistical method to evaluate the set of question banks at College of Economics and Business Administration, Thai Nguyen University.

Key words: Evaluation, reality, question bank, test bank.

Ngày nhận bài: 10/02/2016; Ngày phản biện: 16/03/2016; Ngày duyệt đăng: 31/3/2016

Phản biện khoa học: TS. Nguyễn Quang Hợp – Trường Đại học Kinh tế & Quản trị kinh doanh - DHTN

* Tel: 0912 967 494; Email: Tranccongnghiep@tueha.edu.vn