

ĐÁNH GIÁ ĐỘ TIN CẬY CỦA NGÂN HÀNG ĐỀ THI CHO SINH VIÊN CHUYÊN NGÀNH GIÁO DỤC THỂ CHẤT TRƯỜNG ĐẠI HỌC SƯ PHẠM - ĐẠI HỌC THÁI NGUYÊN

Đào Ngọc Anh*

Tóm tắt:

Độ tin cậy là một tiêu chí quan trọng trong đánh giá một bài thi. Định kì 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, thi thực hành là một bài toán khó. Bài viết đề xuất đánh giá độ tin cậy của bộ ngân hàng câu hỏi thi dành cho sinh viên chuyên ngành Giáo dục thể chất Trường Đại học Sư phạm - Đại học Thái Nguyên theo phương pháp thống kê dựa trên kết quả thi kết thúc học phần.

Từ khóa: Độ tin cậy, ngân hàng câu hỏi thi, ngân hàng đề thi, chuyên ngành Giáo dục thể chất.

Reliability assessment of exam banks based of students majoring in Physical Education at Thai Nguyen University of Education

Summary:

Reliability is an important criterion in evaluating a test. Periodical reliability assessment of testing methods is a requirement of training program accreditation. The reliability assessment of an exam and question banks, which includes multiple exam forms, such as multiple-choice exams, essay exams, oral exams, and practice exams, is a difficult problem. The article proposes to assess the reliability of the bank of exams and questions for students majoring in Physical Education at Thai Nguyen University of Education through statistical method based on the end-of-course examination results.

Keywords: Reliability, bank of exams and questions, Physical Education major.

ĐẶT VẤN ĐỀ

Phân tích và đánh giá các câu hỏi thi - kiểm tra là một việc rất cần thiết nhưng còn ít được quan tâm. Hậu quả là các kì thi sau vẫn thường lặp lại những sai lầm của các kì 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 sinh viên (SV). 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 thiếu rõ ràng, nâng cao 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 SV. Đề thi dễ quá dẫn đến kết quả làm bài của sinh viên quá tốt nên bị xã hội phê phán là chạy theo thành tích, không phản ánh đúng thực lực của người học; Đề khó quá dẫn đến kết quả làm bài của SV kém nên bị xã hội phê phán là chất lượng giáo dục thấp. Tuy nhiên,

*ThS, Trường Đại học Sư phạm - Đại học Thái Nguyên; Email: anhndn@dhsptn.edu.vn

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 có độ khó khác nhau, thậm chí, cùng một đề thi nhưng do những người chấm khác nhau cũng dẫn đến khác nhau.

Cho đến hiện tại, có nhiều công trình nghiên cứu về độ tin cậy của các bài thi trắc nghiệm, tuy vậy, chưa có nghiên cứu nào đề cập đến độ tin cậy của các bài thi tự luận, bài thi vấn đáp, bài thi thực hành. Bằng thống kê chúng tôi nhận thấy, sau 5 lần tổ chức thi đối với các học phần của sinh viên Khoa TĐTT thì có khoảng trên 70% số câu hỏi có trong ngân hàng đề thi đã được sử dụng.

PHƯƠNG PHÁP NGHIÊN CỨU

Trong quá trình nghiên cứu, bài viết đã sử dụng các phương pháp: Phương pháp đọc, phân tích và tổng hợp tài liệu, phương pháp toán học thống kê.

KẾT QUẢ NGHIÊN CỨU VÀ BÀN LUẬN

1. Một số phương pháp kiểm tra đánh giá độ tin cậy của ngân hàng câu hỏi thi

1.1. Phương pháp kiểm tra – kiểm tra lại (Test-Retest Method)

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 gian không quá hai 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 hai 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.

1.2. Phương pháp chia đôi bài kiểm tra (Split Half)

Theo phương pháp này bài kiểm tra được chia thành hai phần bằng nhau, ví dụ một phần toàn câu chẵn, một phần 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 Brown Prophecy 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 Split Half.

r_{AB} : 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_{AB}) có thể dùng công thức:

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

Trong đó:

r_{AB} : 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_{AB} : 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.

1.3. Công thức 20 của Kuder-Richardson (Kuder-Richardson Formula 20)

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

$$R_{tt} = \frac{n}{n-1} \frac{S_t^2 - \sum S_i^2}{S_t^2}$$

Trong đó:

R_{tt} : Độ tin cậy

n: Số câu hỏi trong bài kiểm tra

S_t^2 : Bình phương của độ lệch chuẩn (SD)

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

1.4. Công thức 21 của Kuder-Richardson (Kuder-Richardson Formula 21)

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.

2. Đánh giá độ tin cậy của ngân hàng đề thi

Để đánh giá được phần nào độ tin cậy của các đề thi kết thúc học phần cho sinh viên chuyên ngành GDTC Trường Đại học Sư phạm - Đại học Thái Nguyên trong những năm học gần đây, chúng tôi tiến hành phân tích, đánh giá kết quả thi kết thúc học phần của 2 năm học gần nhất (năm học 2016-2017; 2017-2018) bằng phương pháp thống kê toàn bộ điểm thi kết thúc học phần của tất cả các học phần trong 2 năm học đó và so sánh điểm của các học phần tại các thời điểm khác nhau.

2.1. Phương pháp phân tích

Chúng tôi 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 các học phần đang được giảng dạy cho sinh viên chuyên ngành GDTC Trường Đại học Sư phạm - Đại học Thái Nguyên. Dữ liệu được chọn là toàn bộ kết quả thi kết thúc học phần của 2 năm học

2016-2017 và 2017-2018 trên cơ sở các phiếu điểm lưu trữ tại khoa bao gồm 31 lượt học phần với 528 lượt SV dự thi. Các học phần có tên giống nhau được giảng ở 2 khóa khác nhau làm cơ sở so sánh. Dữ liệu được nhập vào phần mềm Excel, được sử 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 điểm thi ở 2 cận biên của miền điểm là từ 0-1 và 9-10. Các mức đánh giá theo hệ thống tín chỉ tương đương bao gồm F (<4); D (4-5.4); C (5.5-6.9); B (7.0-8.4) 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 khoảng 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 tương đương từ tích D đến tích B, đảm bảo phân loại được SV.

Tổng số 266 lượt học phần, 6.818 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 học phần GDTC có 1.857 bài thi.

Phân tích mức điểm trung bình cao nhất và thấp nhất trong 2 năm học 2016-2017; 2017-2018, kết quả được trình bày ở bảng 1 và bảng 2.

Phân tích theo từng học kỳ, phân điểm trung bình thành 3 khoảng, số liệu cụ thể được trình bày ở bảng 3:

Bảng 1. Bảng tổng hợp số liệu chung của 2 năm học

TT	Năm học	Học kỳ	Lượt môn	Lượt TS dự thi	Hình thức thi (lượt môn)			
					Viết	Vấn đáp	Thực hành	TNMT
1	2016-2017	HK I	48	836	21	10	11	6
2		HKII	51	1178	29	12	5	5
3		HK hè	26	733	12	5	7	3
4	2017-2018	HK I	65	1857	20	7	29	9
5		HKII	41	1235	12	4	22	3
6		HK hè	35	979	12	4	14	5
Tổng			266	6,818				

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

TT	Năm học	Học kì	Điểm trung bình (điểm số)	
			Cao nhất	Thấp nhất
1	2016-2017	HK I	8.84	3.11
2		HK II	8.38	3.46
3	2017-2018	HK I	8.65	3.23
4		HK II	8.35	2.83

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

TT	Năm học	Học kì	Điểm TB từ 0-4.99	Điểm TB từ 5.0-5.59	Điểm từ 5.60-9.0	Tổng
1	2016-2017	HK I	13	6	29	48
2		HKII	15	9	27	51
3		HK hè	5	4	17	26
Tổng			33	19	73	125
4	2017-2018	HK I	12	13	40	65
5		HKII	11	7	23	41
6		HK hè	18	3	14	35
Tổng			41	23	77	141

Qua bảng 3 cho thấy số học phần có mức điểm trung bình từ 5,60 đến 9,0 luôn chiếm tỉ lệ cao so với 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.

2.2. Phân tích theo số bài thi đạt điểm quá thấp hoặc 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.

Nhìn từ bảng trên cho thấy môn Những nguyên lý cơ bản của CN Mác– Lênin; Sinh lý

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

TT	Năm học	Học kỳ	Số bài thi điểm 0-1 và điểm 9-10	
			Thấp nhất 0 - 1	Cao nhất 9 - 10
1	2016 - 2017	HK I	135/328 (41.15%) Tiếng Anh chuyên ngành	31/276 (11.23%) LL& PP TDTT
2		HK II	143/368 (38.85%) Những NLCBCCN Mác - Lênin; 30/364 (8.24%) Y học TDTT	76/363 (20,9%) Lý thuyết XSTK
3	2017 - 2018	HK I	131/414 (31.6%) Đo lường TDTT	174/345 (50.43%) Lịch sử TDTT
4		HK II	129/277 (46.57%) Giáo dục học	22/216 (10.18%) Tâm lý TDTT

TĐTT; Đo lường TĐTT và Giáo dục học có số bài thi đạt điểm 0-1 cao và có tính lặp lại ở các năm. Môn Lịch sử TĐTT có điểm đạt 9-10 với số bài quá nhiều 174/345 (50.43%).

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, cụ thể như sau:

Bảng 5. Bảng phân tích độ lệch chuẩn

TT	Năm học	Học kỳ	ĐLC từ 0-1.49	Tỉ lệ %	ĐLC từ 1.5-1.8	Tỉ lệ %	ĐLC từ 1.81 trở lên	Tỉ lệ %
1	2016-2017	HK I	29	36.25	23	38.99	50	37.59
2		HK II	35	43.75	28	47.46	46	34.59
3		HK hè	16	20	8	13.55	37	27.82
Tổng			80		59		133	
4	2017-2018	HK I	45	40.91	28	41.19	66	50.77
5		HK II	37	33.64	21	30.88	30	23.07
6		HK hè	28	25.45	16	23.53	34	26.16
Tổng			110		68		130	

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 chiếm tỉ 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. Kết quả phân tích cho thấy có sự chênh lệch lớn về giá trị của từng thời điểm đánh giá so với giá trị trung bình.

KẾT LUẬN

Độ tin cậy là điều kiện cần của các đề thi 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 hay tăng độ phân biệt của các câu hỏi. 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 2 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.

TÀI LIỆU THAM KHẢO

1. Nguyễn Công Khanh (2004), *Đánh giá và đo lường trong KHXH*, Nxb chính trị QG, HN.
2. Henning, G. (1987), *A Guide to Language Testing*, Cambridge, Newbury House Publishers.
3. Heaton, J.B. (1988), *Writing English Language Tests*, London: Long Man.

4. Bachman, L.F (1990), *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press.

5. Thorndike&Haghen, *Đo lường và đánh giá trong tâm lý và giáo dục*, bản dịch của ĐHQGHN.

(Bài nộp ngày 27/11/2020, phản biện ngày 2/1/2021, duyệt in ngày 21/4/2021)