

DOI: 10.58490/ctump.2025i92.4236

**ỨNG DỤNG LÝ THUYẾT TRẮC NGHIỆM CỔ ĐIỂN TRONG ĐÁNH GIÁ
ĐỀ THI TỐT NGHIỆP BÁC SĨ NGÀNH Y KHOA CỦA KHOA Y –
TRƯỜNG ĐẠI HỌC NGUYỄN TẮT THÀNH NĂM HỌC 2024 – 2025**

Nguyễn Ngọc Khánh, Lê Thị Kim Nhung, Lê Minh Nguyệt, Phạm Tú Quyên*

Trường Đại học Nguyễn Tấn Thành

**Email: nnkhanh@ntt.edu.vn*

Ngày nhận bài: 14/9/2025

Ngày phản biện: 10/10/2025

Ngày duyệt đăng: 25/10/2025

TÓM TẮT

Đặt vấn đề: Đánh giá chất lượng đề thi là một phần quan trọng của công tác đảm bảo chất lượng giáo dục. **Mục tiêu nghiên cứu:** Đánh giá chất lượng của bộ đề thi tốt nghiệp bác sĩ ngành Y khoa của khoa Y – Nguyễn Tấn Thành năm học 2024 – 2025. **Đối tượng và phương pháp nghiên cứu:** Hồi cứu kết quả bài thi và sử dụng các chỉ số trong lý thuyết trắc nghiệm cổ điển để phân tích chất lượng của từng câu hỏi và chất lượng của đề thi. **Kết quả:** Đề thi có độ tin cậy và độ giá trị tốt. Các yếu tố ảnh hưởng đến chất lượng của đề thi bao gồm độ khó, độ phân biệt, r -pbis của từng câu hỏi thi, số lượng câu hỏi và thời gian làm bài thi. **Kết luận:** Đề thi có độ giá trị vì đạt được mục tiêu của người ra đề. Phổ điểm của sinh viên tập trung xung quanh giá trị điểm thi trung bình. Tuy nhiên chưa thể đánh giá năng lực của thí sinh một cách chính xác.

Từ khóa: Độ khó, độ tin cậy, độ phân biệt, lý thuyết trắc nghiệm cổ điển, câu hỏi trắc nghiệm khách quan.

ABSTRACT

**USING CLASSICAL TEST THEORY TO EVALUATE
THE GRADUATION EXAM FOR GENERAL DOCTOR OF
THE MEDICINE FACULTY – NGUYEN TAT THANH UNIVERSITY
IN 2024 – 2025**

Nguyen Ngoc Khanh, Le Thi Kim Nhung, Le Minh Nguyệt, Pham Tu Quyên*

Nguyen Tat Thanh University

Background: Evaluating the quality of the exam is an important part of Education Quality Assurance. **Objectives:** To evaluate the quality of the graduation exam for general doctor of the Medicine Faculty – Nguyen Tat Thanh University in 2024-2025. **Materials and methods:** Retrospecting the results of the exam and using classical test theory to analyze the quality of each multiple-choice question (MCQ) and the quality of the exam. **Results:** The exam has good reliability and validity. Factors affecting the quality of the exam include difficulty index, discrimination index, r -pbis of each exam question, quantity of MCQs and duration of the exam. **Conclusions:** The examination demonstrates content validity in that it meets the objectives set by the test designer. The score distribution is concentrated around the mean examination score. However, it does not yet allow for an accurate assessment of candidates' competencies.

Keywords: Difficulty index, reliability index, discrimination index, classical test theory, multiple-choice question (MCQ).

I. ĐẶT VẤN ĐỀ

Kỳ thi tốt nghiệp là kỳ thi quan trọng nhất trong 6 năm học bác sĩ ngành Y khoa. Trong đó bài thi trắc nghiệm lý thuyết giúp đánh giá chính xác kiến thức tổng quát các

chuyên khoa của sinh viên. Đề thi tốt nghiệp phải đảm bảo đúng về chuyên môn, kỹ thuật và quy trình ra đề thi. Việc phân tích kết quả thi giúp chúng ta đánh giá chính xác về chất lượng đề thi (số lượng câu hỏi có đủ nhiều để đảm bảo đánh giá chính xác trình độ của sinh viên, chất lượng của từng câu hỏi thi có tốt không và đề thi có thể phân loại được sinh viên hay không...) để từ đó chúng ta sẽ có sự điều chỉnh về nội dung, phương pháp dạy học nhằm nâng cao chất lượng giáo dục [1]. Vì vậy nghiên cứu được thực hiện với mục tiêu: 1) Đánh giá đề thi tốt nghiệp bác sĩ ngành Y khoa – Trường Đại học Nguyễn Tất Thành (NTTU) năm học 2024 – 2025; 2) Đánh giá chất lượng câu hỏi trong bài thi tốt nghiệp bác sĩ ngành Y khoa – NTTU năm học 2024 – 2025; 3) Hệ thống hóa các tiêu chí cần thiết để đánh giá chất lượng bộ đề thi và câu hỏi thi.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Thiết kế nghiên cứu: Mô tả cắt ngang, hồi cứu kết quả bài thi tốt nghiệp của sinh viên Y6 – NTTU năm học 2024 – 2025 (bao gồm 120 câu và 118 sinh viên tham gia thi). Sử dụng phần mềm Excel và Epicalc để xử lý số liệu. Sử dụng các chỉ số: độ khó, độ khó kỳ vọng, độ phân biệt, r-pbis, độ tin cậy và độ giá trị để đánh giá chất lượng đề thi và chất lượng câu hỏi thi.

2.1. Độ khó (chỉ số độ khó): Difficulty Index (Ký hiệu là P)

Độ khó hay chỉ số độ khó của câu hỏi trắc nghiệm khách quan (MCQ) là khả năng chọn đúng đáp án chính xác của câu hỏi đó (“chỉ số độ khó” chỉ thể hiện khả năng số lượng thí sinh trả lời đúng câu hỏi đó nhiều hay ít, tuy nhiên điều này không phản ánh được là kiến thức của câu hỏi đó là khó hay dễ so với năng lực của thí sinh). Do đó “độ khó” hay chỉ số độ khó càng cao thì câu hỏi đó là dễ (vì có nhiều người trả lời đúng) và ngược lại “độ khó” càng thấp thì câu hỏi càng khó (vì có nhiều người trả lời sai).

Theo L.Q. Thiệp [1] và N.T. Hiếu [2] thì công thức tính độ khó là:

$$P_i = n_i / N_i$$

P_i : Độ khó của câu hỏi thứ i trong bộ đề thi.

n_i : Số thí sinh trả lời đúng câu hỏi thứ i trong bộ đề thi.

N_i : Số thí sinh có trả lời câu hỏi thứ i trong bộ đề thi.

Giá trị độ khó của mỗi câu hỏi có ý nghĩa hay không còn phụ thuộc vào số lượng thí sinh tham gia trả lời câu hỏi đó. Ví dụ 2 câu hỏi có cùng “độ khó” như 10/20 và 30/60 thì rõ ràng câu hỏi có độ khó là 30/60 sẽ phản ánh chính xác hơn do có mẫu lớn hơn (do đó cần so sánh với độ khó kỳ vọng của câu hỏi đó và số lượng thí sinh kỳ vọng là số lượng thí sinh tham gia thi).

Độ khó trung bình của đề thi ($P.tb$) là trung bình độ khó của tất cả các câu hỏi trong bộ đề thi đó.

$$P.tb = (\sum P_i / \text{tổng số câu hỏi}) \pm SD \quad (SD: \text{độ lệch chuẩn của trung bình độ khó})$$

2.2. Độ khó kỳ vọng: Expected Difficulty Index (Ký hiệu là P.kv)

Độ khó kỳ vọng của một câu hỏi trắc nghiệm khách quan phụ thuộc vào số lượng đáp án của câu hỏi đó. Như vậy số lượng đáp án càng nhiều thì độ khó kỳ vọng càng cao.

Theo L.Q.Thiệp [1] thì công thức tính độ khó kỳ vọng là:

$$P.kv = (100\% + 1/Y)/2 \quad (Y: \text{số đáp án của câu hỏi trắc nghiệm})$$

Nếu $Y = 4$ thì $P.kv = 0,625$

Độ khó kỳ vọng trung bình của đề thi là trung bình độ khó kỳ vọng của từng câu hỏi trong bộ đề thi đó. Vì vậy đối với bộ đề thi mà các câu hỏi có chung số lượng đáp án thì độ

khó kỳ vọng của đề thi bằng với độ khó kỳ vọng của một câu hỏi trong đề thi đó. Và vì độ khó kỳ vọng của các câu hỏi là như nhau nên độ lệch chuẩn $SD = 0$. Như vậy độ khó kỳ vọng trung bình của 1 bộ đề thi (gồm các câu hỏi MCQ có 4 đáp án) sẽ là: 0,625.

2.3. Độ phân biệt: Discrimination Index (Ký hiệu là D)

Từ kết quả thi chúng ta chia ra 2 nhóm là nhóm điểm cao (27% thí sinh có điểm cao nhất) và nhóm điểm thấp (27% thí sinh có điểm thấp nhất). Độ phân biệt là khả năng nhóm điểm cao trả lời được câu hỏi MCQ đó cao hơn hay thấp hơn so với nhóm điểm thấp. Như vậy câu hỏi có độ phân biệt tốt là câu hỏi mà số lượng thí sinh của nhóm điểm cao phải trả lời được nhiều hơn nhóm điểm thấp và đề thi có nhiều câu hỏi có độ phân biệt tốt sẽ giúp phân biệt được các nhóm thí sinh có năng lực giỏi, khá, trung bình hay yếu kém.

Theo L.Q. Thiệp [1] thì công thức tính độ phân biệt là:

$$D_i = (C_i - T_i)/S$$

D_i : độ phân biệt của câu hỏi thứ i trong bộ đề thi.

C_i : Số lượng thí sinh trong nhóm điểm cao trả lời đúng câu hỏi thứ i .

T_i : Số lượng thí sinh trong nhóm điểm thấp trả lời đúng câu hỏi thứ i .

S : là 27% số lượng thí sinh của tổng số thí sinh.

Một điều lưu ý khi tính toán độ phân biệt là phải chú ý đến giá trị P-value. Nếu P-value $\leq 0,05$ thì sự khác biệt của nhóm điểm cao và nhóm điểm thấp mới có ý nghĩa. Nếu P-value $> 0,05$ thì không có giá trị phân biệt giữa hai nhóm. Do đó số lượng thí sinh phải đủ nhiều và sự khác biệt phải đủ lớn thì mới biết chính xác câu hỏi có độ phân biệt tốt hay không.

2.4. Chỉ số r-pbis (point biserial correlation)

Hệ số tương quan điểm nhị phân r-pbis của từng câu hỏi MCQ thể hiện mối tương quan giữa khả năng trả lời đúng câu hỏi đó và điểm số của thí sinh (được tính và giải thích tương tự như hệ số tương quan Pearson) [1], [3].

Như vậy chỉ số r-pbis cũng tương ứng với độ phân biệt (D) tuy nhiên r-pbis tính toán trên tổng số thí sinh trong khi đó độ phân biệt (D) thì tính toán trên 54% số thí sinh. Và khi phân tích chỉ số r-pbis của từng câu hỏi thì cũng phải chú ý đến giá trị của P-value. Nếu P-value $\leq 0,05$ thì giá trị thu được là có ý nghĩa thống kê. Nếu P-value $> 0,05$ thì giá trị này có thể là ngẫu nhiên.

2.5. Độ tin cậy (Reliability) ký hiệu là r (Kuder21) hoặc r (Kuder20)

Đề thi trắc nghiệm là một phép đo năng lực của thí sinh do đó cần phải tính độ tin cậy để biết được phép đo đó có chính xác hay không. Để tính độ tin cậy người ta có thể dựa vào phương pháp Kuder-Richardson [1].

$$r_{20} = \left[\frac{K}{K-1} \right] \left[1 - \frac{\sum P_i Q_i}{\sigma^2} \right] \quad \text{Hay} \quad r_{21} = \left[\frac{K}{K-1} \right] \left[1 - \frac{\mu(K-\mu)}{K\sigma^2} \right]$$

r_{20} : r (Kuder 20): áp dụng khi các câu hỏi có độ khó khác nhau. Nếu thiếu dữ liệu độ khó của 1 số câu hỏi thì r_{20} sẽ không phản ánh chính xác độ tin cậy của bộ đề thi.

r_{21} : r (Kuder 21): áp dụng khi r_{20} không tính được do thiếu dữ liệu độ khó của 1 số câu hỏi.

K : tổng số câu hỏi trong đề thi

P_i : độ khó của câu hỏi

$Q_i = (1 - P_i)$: tỉ lệ trả lời sai của câu hỏi thứ i

σ^2 : phương sai của điểm thi trung bình đề thi

μ : điểm thi trung bình của đề thi

Qua công thức trên chúng ta thấy rằng số lượng câu hỏi càng nhiều thì độ tin cậy của bộ đề thi càng cao. Ngoài ra độ tin cậy còn phụ thuộc vào điểm trung bình của thí sinh, độ khó của câu hỏi và số lượng câu hỏi được trả lời.

2.6. Độ giá trị (Validity)

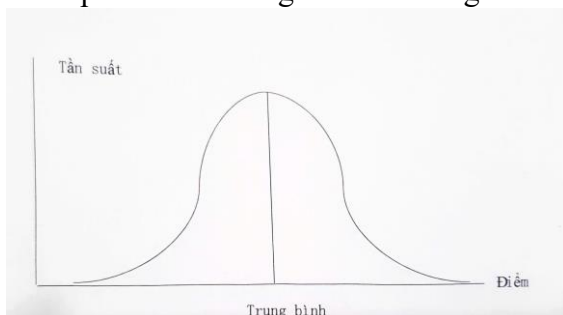
Độ giá trị của một bộ đề thi trắc nghiệm là khả năng đánh giá được năng lực của thí sinh. Khi thiết kế bộ đề thi trắc nghiệm chúng ta sẽ cân nhắc xem đề thi cần có bao nhiêu câu hỏi, câu hỏi có bao nhiêu đáp án và thời gian làm bài là bao lâu để có thể bao quát hết nội dung kiến thức và đánh giá được năng lực của thí sinh (phân loại được thí sinh giỏi, khá, trung bình và yếu kém).

Theo L. Q. Thiệp [1] thì có thể so sánh điểm thi trung bình thực tế với điểm thi trung bình kỳ vọng của đề thi đó. Nếu kết quả là tương đồng thì đề thi đánh giá được khả năng của thí sinh theo đúng mong muốn của người ra đề. Tuy nhiên để có thể đánh giá kết quả thi có phản ánh đúng năng lực của thí sinh hay không thì nên phân tích thêm sự tương quan giữa điểm thi của bộ đề thi và kết quả học tập của thí sinh.

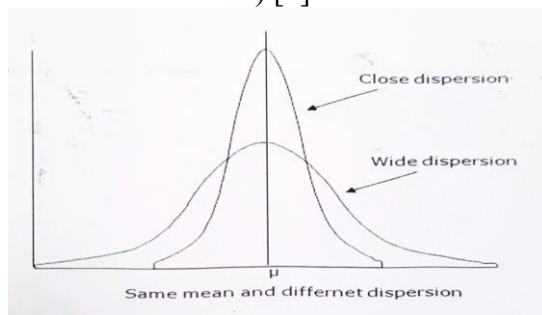
Điểm thi trung bình kỳ vọng = $P.kv \times 10$ (đánh giá theo thang điểm 10)

Lưu ý: các câu hỏi trong đề thi có cùng số lượng đáp án.

Ngoài ra để phân tích xem đề thi có phân tách được các nhóm sinh viên có trình độ khác nhau hay không thì cần chú ý xem phổ điểm thi phân tán hay đồng đều, có phân bố chuẩn hay không (đường cong hình chuông, đa số điểm thi sẽ tập trung quanh điểm thi trung bình và các phổ điểm ở càng xa điểm trung bình thì sẽ có ít thí sinh hơn) [4].



Hình 1. Phân bố chuẩn của phổ điểm



Hình 2. Phân bố có độ phân tán rộng và hẹp

III. KẾT QUẢ NGHIÊN CỨU

3.1. Phân tích câu hỏi thi

Bảng 1. Thông số phân tích của 120 câu hỏi

Câu	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
P	0,14	0,24	0,47	0,73	0,32	0,31	0,75	0,92	0,69	0,86	0,9	0,97	0,92	0,43	0,9
D	0,19	0,19	0,22	0,44	0,31	0,25	0,22	0,16	0,22	0,22	0,13	0,09	0,13	0,16	0,19
r-pbis	0,16	0,25	0,22	0,41	0,26	0,21	0,2	0,3	0,21	0,34	0,2	0,16	0,1	0,07	0,28
Câu	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
P	0,61	0,45	0,92	0,93	0,92	0,45	0,74	0,56	0,9	0,96	0,72	0,97	0,76	0,86	0,83
D	0,19	0,22	0,09	0,06	0,16	0,28	0,06	0,13	0,22	0,13	0,34	0,06	0,19	0,09	0,22
r-pbis	0,2	0,13	0,12	0,11	0,24	0,21	0,09	0,15	0,34	0,32	0,31	0,15	0,12	0,1	0,25
Câu	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
P	0,88	0,88	0,61	0,42	0,6	0,5	0,3	0,25	0,59	0,99	0,7	0,55	0,58	0,21	0,84
D	0,16	0,16	0,09	0,22	0,13	0,47	0	<0	0,38	0,03	0,09	<0	0,16	0	0,09
r-pbis	0,25	0,14	0,08	0,1	<0	0,32	<0	<0	0,29	0,09	0,07	<0	0,18	0,08	0,05

Câu	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
P	0,9	0,47	0,71	0,97	0,47	0,76	0,36	0,94	0,25	0,75	0,33	0,47	0,46	0,86	0,89
D	0,09	0,19	0,38	0,03	0,19	0,16	<0	0,03	0,06	0,09	0,22	0,34	0,31	0,25	0,16
r-pbis	0,13	0,14	0,23	0,02	0,06	0,17	<0	0,07	0,07	0,14	0,18	0,26	0,25	0,3	0,22
Câu	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
P	0,98	0,53	0,59	0,85	0,41	0,99	0,33	0,88	0,98	0,7	0,53	0,58	0,64	0,94	0,95
D	0,03	0,44	0,22	0,31	0,13	0	0,09	0	0,03	0,38	0,22	0,13	0,34	<0	0
r-pbis	0,08	0,3	0,13	0,32	0,13	0,02	0,17	0,08	0,16	0,26	0,21	0,15	0,32	<0	<0
Câu	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
P	0,92	0,91	0,53	0,81	0,68	0,85	0,9	0,75	0,35	0,42	0,59	0,74	0,83	0,83	0,85
D	<0	0,16	0,43	0,31	0,47	0,19	0,13	0,25	0,19	0,22	0,34	0,13	<0	0,28	0,34
r-pbis	0,01	0,25	0,29	0,28	0,42	0,3	0,14	0,24	0,16	0,16	0,31	0,14	<0	0,3	0,37
Câu	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
P	0,91	0,75	0,77	0,97	1	1	0,98	1	0,99	0,89	0,81	0,39	0,91	0,47	0,9
D	0,06	0,19	0	0	0	0	0,03	0	0,03	0,16	0,06	0,34	0,28	0,25	0,19
r-pbis	0,12	0,15	0,1	<0			0,1		0,13	0,22	0,12	0,31	0,41	0,17	0,29
Câu	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
P	0,96	0,64	0,32	0,88	0,94	0,36	0,66	0,49	0,63	0,48	0,39	0,92	0,28	0,48	0,25
D	0,06	0,44	<0	0,31	0	0,28	0,34	0,06	0,03	0,03	<0	0,13	<0	0,13	0,38
r-pbis	0,15	0,35	0,07	0,4	0,09	0,26	0,37	0,11	0,05	0,05	<0	0,17	0,01	0,05	0,31

Nhận xét: 118 thí sinh trả lời tất cả 120 câu hỏi (không có câu hỏi nào bị bỏ trống).
P-value của D: đa số > 0,05. P-value của r-pbis: tất cả > 0,05.

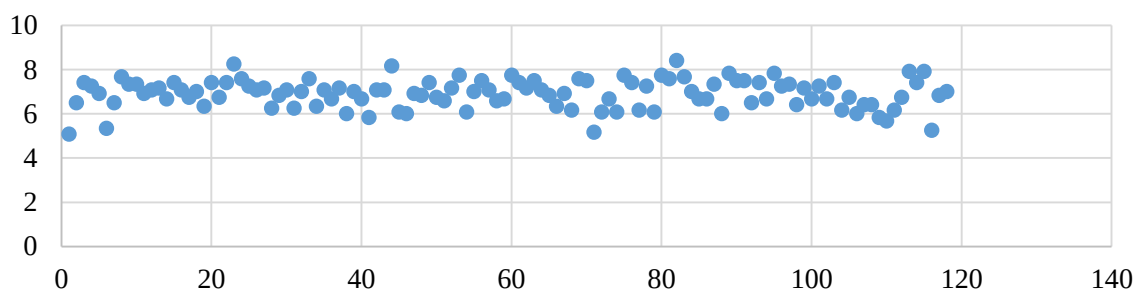
3.2. Phân tích độ tin cậy của đề thi

Bảng 2. Thông số phân tích của bộ đề thi

	Giá trị	P-value
P.tb của đề thi	0,69 ± 0,24	< 0,05
P.kv của đề thi	0,625	
Điểm trung bình thực tế của đề thi	6,91 ± 0,65	< 0,05
Điểm trung bình kỳ vọng của đề thi	6,25	
r (Kuder21)	0,9	
r (Kuder20)	0,7	

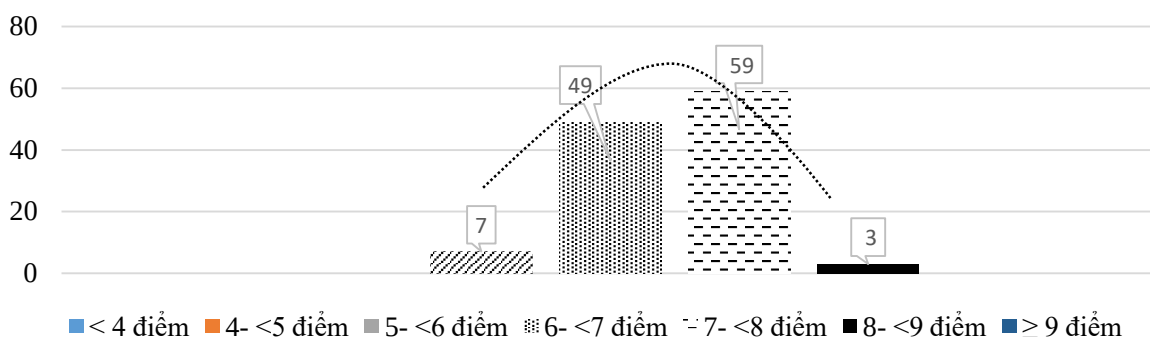
Nhận xét: Điểm thi trung bình thực tế của đề thi cao hơn so với điểm thi trung bình kỳ vọng.

3.3. Phân tích độ giá trị của đề thi



Biểu đồ 1. Phân bố điểm thi của 118 sinh viên

Nhận xét: Điểm thi của thí sinh chủ yếu tập trung ở phổ từ 6 điểm đến 8 điểm.



Biểu đồ 2. Phân bố theo nhóm điểm

Nhận xét: Nhóm (7- <8 điểm) chiếm tỉ lệ lớn nhất (50%).

IV. BÀN LUẬN

Qua việc phân tích kết quả của kỳ thi chúng tôi nhận thấy độ khó trung bình cao hơn độ khó kỳ vọng của đề thi (sự khác biệt có ý nghĩa thống kê với $P\text{-value} < 0,05$). Điều này có nghĩa là bộ đề thi có mức độ dễ hơn so với năng lực của thí sinh. Tác giả V. C. Danh [5] cũng có quan điểm giống chúng tôi. Đi vào phân tích cụ thể từng câu hỏi, chúng tôi nhận thấy chất lượng của từng câu hỏi không đồng đều với nhau (về độ khó, độ phân biệt và $r\text{-pbis}$).

Xét về độ khó của từng câu hỏi. Theo L. Q. Thiệp thì P nằm trong khoảng 0,25 – 0,75 là độ khó chấp nhận được, nếu $P < 0,25$ là câu hỏi quá khó và $P > 0,75$ là quá dễ [1]. Còn theo Osterlind thì độ khó trong khoảng 0,4 – 0,8 là chấp nhận được [6]. Đề thi của chúng tôi có 3 câu hỏi (1-2-44) ở mức độ quá khó ($P < 0,25$), có 15 câu hỏi (5-6-37-38-52-54-56-67-84-102-108-111-116-118-120) ở mức độ khó (0,25 - <0,4), có 46 câu hỏi ở mức độ trung bình (0,4 – 0,75) và có 56 câu hỏi ở mức độ quá dễ ($P > 0,75$) (đặc biệt câu 95-96-98 có 100% thí sinh trả lời đúng). Việc số lượng câu hỏi quá dễ chiếm tỷ trọng lớn trong bộ đề thi đã đưa đến tình trạng đề thi dễ hơn năng lực của sinh viên. Để đánh giá chất lượng của một câu hỏi MCQ còn cần phải dựa vào độ phân biệt D và chỉ số tương quan $r\text{-pbis}$. Nếu D và/hoặc $r\text{-pbis} \leq 0$ thì là câu hỏi có vấn đề (nhóm điểm thấp có nhiều khả năng trả lời đúng câu hỏi đó hơn nhóm điểm cao). Điều này có thể được lý giải như sau: những câu hỏi đó có thể quá khó (nên thí sinh giỏi cũng không trả lời được nên đa số thí sinh sẽ chọn đáp án dựa vào sự may rủi chứ không phải dựa vào kiến thức) hoặc có thể là những câu hỏi quá dễ (hầu như thí sinh nào cũng trả lời được nên không có sự khác biệt giữa nhóm điểm cao và nhóm điểm thấp). Với D và $r\text{-pbis} > 0$ thì theo tác giả L. Q. Thiệp hai giá trị này càng lớn và càng tiến gần đến giá trị +1 thì câu hỏi có độ phân cách càng tốt và càng có khả năng phân biệt năng lực của thí sinh (D và $r\text{-pbis} \geq 0,2$ là chấp nhận được) [1]. Tuy nhiên khi xét D và $r\text{-pbis}$ của một câu hỏi MCQ thì cần chú ý đến giá trị $P\text{-value}$. Nếu $P\text{-value} \leq 0,05$ thì kết quả đó mới có ý nghĩa và chính xác, còn nếu $P\text{-value} > 0,05$ thì kết quả đó có thể chỉ là ngẫu nhiên và cần phải xem xét kỹ trước khi đưa ra nhận định.

Như vậy một câu hỏi MCQ tốt phải đảm bảo có độ khó ở mức độ trung bình, chỉ số D và $r\text{-pbis}$ ở mức độ chấp nhận được và $P\text{-value} \leq 0,05$. Trong bộ đề thi của chúng tôi chỉ có 18 câu là tương đối tốt tuy nhiên $P\text{-value} > 0,05$ có nghĩa là kết quả thu được chưa hoàn toàn chính xác. Qua đó có thể thấy rằng để tạo ra một bộ đề thi trắc nghiệm mà tất cả các câu hỏi đều tốt là rất khó. Và trong số những câu hỏi chưa tốt thì nội dung câu hỏi đó có vấn đề gì hay không? Ví dụ những câu hỏi có độ khó cao (nghĩa là quá dễ nên đa số thí

sinh đều làm được). Vậy nếu thí sinh không có kiến thức có khả năng làm đúng những câu hỏi quá dễ này hay không. Tỷ lệ để 1 thí sinh (không có kiến thức) làm đúng 1 câu hỏi MCQ có 4 đáp án dựa vào may rủi là 25% nhưng để 118 thí sinh (không có kiến thức) cùng làm đúng dựa vào may rủi là $(0,25)^{118}$ (tỷ lệ này gần như là bằng 0). Vậy trong trường hợp này có thể kết luận rằng những câu hỏi mà đa số thí sinh làm được là những câu có kiến thức cơ bản, nội dung câu hỏi rõ ràng (không đánh đố) hoặc thí sinh ôn tập đúng nội dung. Tuy nhiên để làm được những câu hỏi quá dễ thì cũng đòi hỏi thí sinh có nền tảng kiến thức nhất định. Đối với những câu hỏi quá khó thì vấn đề là kiến thức quá khó, thí sinh không học kiến thức đó hoặc lỗi kỹ thuật khi đặt câu hỏi. Nếu vấn đề là do thí sinh thì cần tìm hiểu lý do tại sao. Nếu do lỗi kỹ thuật thì nên sửa lỗi nếu có thể hoặc loại bỏ câu hỏi ra khỏi đề thi. Nếu do kiến thức quá khó thì vẫn có thể để lại trong đề thi với tỷ trọng phù hợp để tìm kiếm thí sinh có thể đạt điểm 10 một cách xứng đáng. Chúng tôi cho rằng bộ đề thi có thể có tối đa 5 câu hỏi có kiến thức rất khó vì xác suất trả lời đúng 5 câu hỏi MCQ 4 đáp án theo cách may rủi là $(0,25)^5 \approx 1/1000$. Như vậy một thí sinh có thể đạt được điểm 10 chắc chắn phải có kiến thức xuất sắc chứ không thể chỉ dựa vào may mắn hoặc là bị lộ đề.

Theo L. Q. Thiệp thì đề có bộ đề thi tốt thì nên có nhiều câu hỏi có độ khó trung bình, độ phân cách và r-pbis ở mức chấp nhận được [1]. Tuy nhiên chúng ta có thể thấy điều đó là rất khó và những câu hỏi có độ khó cao hoặc thấp cũng có giá trị nhất định trong bộ đề thi. Như vậy để đánh giá chất lượng của bộ đề thi trắc nghiệm thì cần phải chú ý đến độ tin cậy và độ giá trị của nó (nghĩa là kết quả của bộ đề thi có phản ánh chính xác năng lực của thí sinh và có phân loại được trình độ của thí sinh hay không). Theo công thức tính r (kuder 20) và r (Kuder 21) thì độ tin cậy của bộ đề thi phụ thuộc vào số lượng câu hỏi thi. Số lượng câu hỏi thi càng nhiều thì độ tin cậy càng cao. Chúng tôi thấy rằng số lượng câu hỏi thi càng nhiều thì càng bao quát được nhiều nội dung kiến thức của sinh viên (tránh tình trạng học tủ, học lệch...). Ngoài ra độ tin cậy cũng phụ thuộc vào số lượng câu hỏi được thí sinh trả lời (điều này phụ thuộc vào thời gian của đề thi. Nếu cho ít thời gian quá thì thí sinh sẽ không làm kịp và có thể bỏ trống nhiều câu hỏi hoặc là họ sẽ đánh theo kiểu may rủi dẫn đến giảm độ tin cậy của bộ đề thi). Theo Brennan [7] và D. T. Anh [8] thì độ tin cậy có giá trị $\geq 0,7$ là tốt. Kết quả phân tích cho thấy đề thi của chúng tôi có độ tin cậy tốt với r (Kuder 20) là 0,7 và r (Kuder 21) là 0,9. Độ giá trị của đề thi nghĩa là bộ đề thi có khả năng đạt được mục tiêu của người ra đề hay không. Khi làm ra bộ đề thi (gồm K số câu hỏi với Y đáp án) thì người ra đề đã đặt ra một số điểm trung bình kỳ vọng tương ứng và nếu kết quả thu được (điểm thi trung bình thực tế) tương đương hoặc cao hơn điểm trung bình kỳ vọng thì có nghĩa là đề thi có giá trị giúp đánh giá năng lực người học. Theo kết quả ở bảng 2 thì bộ đề thi của chúng tôi có độ giá trị. Tuy nhiên người ta còn kỳ vọng bộ đề thi sẽ giúp phân chia thí sinh vào các nhóm các trình độ năng lực khác nhau như giỏi (8- <9 điểm, ≥ 9 điểm), khá (6- <7 điểm, 7- <8 điểm), trung bình (4- <5 điểm, 5- <6 điểm) và yếu kém (<4 điểm). Theo kết quả nghiên cứu của chúng tôi (biểu đồ 1 và 2) thì phổ điểm chủ yếu tập trung xung quanh giá trị điểm thi trung bình và ở 2 nhóm là (6- <7 điểm) và (7- <8 điểm) chiếm 91.5%. Như vậy độ phân tán của điểm thi là hẹp. Có tình trạng này là do độ phân cách của các câu hỏi trong đề thi là chưa được tốt. Tuy nhiên cũng có thể là do năng lực và trình độ của 118 thí sinh tương đối đồng đều nên có thể dẫn đến kết quả trên. Theo L. Q. Thiệp thì để biết bộ đề thi có đánh giá chính xác trình độ của thí sinh hay không thì cần phải tính hệ số tương quan giữa điểm thi và điểm trung bình kết quả học tập 6 năm của mỗi thí sinh. Nếu tương quan thuận và chặt chẽ thì đề thi có giá trị tốt [1]. Đáng tiếc là chúng tôi không có được điểm kết quả học tập 6 năm của 118 thí sinh nên không thể tính được hệ số tương quan này.

V. KẾT LUẬN

Đề thi dễ hơn so với năng lực của thí sinh (độ khó trung bình cao hơn độ khó kỳ vọng của đề thi). Đề thi có độ tin cậy tốt (số lượng câu hỏi và thời gian thi phù hợp). Đề thi có độ giá trị vì đạt được mục tiêu của người ra đề (điểm thi trung bình thực tế cao hơn điểm thi trung bình kỳ vọng của bộ đề thi). Phổ điểm của sinh viên tập trung xung quanh giá trị điểm thi trung bình. Tuy nhiên chưa thể đánh giá năng lực của thí sinh một cách chính xác. Do đó chúng tôi kiến nghị như sau: Cần chú ý cải thiện độ khó, độ phân biệt và chỉ số r -pbis của các câu hỏi trong bộ đề thi. Cần đánh giá mối tương quan giữa kết quả đề thi tốt nghiệp với điểm trung bình 6 năm học để tìm hiểu xem đề thi tốt nghiệp có phản ánh chính xác trình độ năng lực thực sự của thí sinh.

TÀI LIỆU THAM KHẢO

1. Lâm Quang Thiệp. Đo lường trong giáo dục lý thuyết và ứng dụng, NXB Đại học quốc gia, Hà Nội. 2010. trang 51-71, <https://www.slideshare.net/slideshow/trc-nghim-v-o-lng-trong-gio-dc-gstskh-lm-quang-thip/250784161>.
 2. Nguyễn Trung Hiếu, Nguyễn Bích Như. Ứng dụng lý thuyết trắc nghiệm cổ điển trong phân tích câu hỏi trắc nghiệm khách quan. *Tạp chí Khoa học trường đại học Đồng Tháp*. 2018. (35), 111-116, DOI: 10.52714/dthu.35.12.2018.654.
 3. Bùi Anh Tú, Võ Đăng Khoa và cộng sự. Đánh giá đề thi MCQ đầu vào và tương quan độ khó, độ phân cách các câu, *Tạp chí Y học thành phố*. 2024. 27(5), 38-47, DOI 10.32895/hcjm.m.2024.05.05.
 4. Hessing, T. 2025. Dispersion, <https://sixsigmastudyguide.com/dispersion>.
 5. Võ Công Danh, Bùi Ngọc Quang. Xây dựng và đánh giá câu hỏi trắc nghiệm khách quan, *Tạp chí Khoa học trường đại học Đồng Tháp*. 2016. (23), 16-21, DOI: 10.52714/dthu.23.12.2016.409.
 6. Osterlind, S. J. Constructing test items, *Boston: Kluwer Academic*. 1989. (15), 35-47, <http://dx.doi.org/10.1007/978-94-009-1071-3>.
 7. Brennan, L. R. Generalizability Theory and Classical Test Theory, *Applied Measurement in Education*, 2010. (24), 1-21, <https://doi.org/10.1080/08957347.2011.532417>.
 8. Dương Thị Anh, Nguyễn Thu Hà. Độ giá trị đề thi và việc đảm bảo độ giá trị đề thi trong các kỳ thi diện rộng ở Việt Nam. 2019. <http://qm.education.vnu.edu.vn/do-gia-tri-de-thi>.
-