

PHÂN TÍCH CÂU TRẮC NGHIỆM CỦA CÁC BÀI THI TUYỂN SINH CHƯƠNG TRÌNH CHUYÊN KHOA CẤP I GIẢI PHẪU HỌC VÀ SINH LÝ HỌC TẠI ĐẠI HỌC Y DƯỢC THÀNH PHỐ HỒ CHÍ MINH TỪ NĂM 2018 ĐẾN 2023 THEO LÝ THUYẾT TRẮC NGHIỆM CỔ ĐIỂN

Trần Thị Diệu¹, Phạm Dương Uyên Bình², Bùi Trí Dũng³, Mai Phương Thảo³, Nguyễn Anh Vũ⁴,
Vĩnh Sơn², Phạm Lê An⁵

TÓM TẮT

Mục tiêu: Nghiên cứu để đánh giá độ khó của câu hỏi, độ phân cách D 27%, độ phân cách r_{pbis} và độ tin cậy của các đề thi trắc nghiệm câu hỏi nhiều lựa chọn (MCQs) trong kỳ thi tuyển sinh cho chương trình Chuyên khoa cấp I đối với hai môn thi cơ sở là Giải phẫu học và Sinh lý học qua các năm nhằm nâng cao chất lượng đề thi và câu hỏi thi.

Đối tượng và Phương pháp nghiên cứu: Nghiên cứu cắt ngang được thực hiện trên đề thi tuyển đầu vào chương trình đào tạo Chuyên khoa cấp I tại Đại học Y Dược TP. Hồ Chí Minh từ năm 2018 đến năm 2023 với hai môn Giải phẫu học và Sinh lý học. Mỗi đề thi có 120 câu hỏi trắc nghiệm 4 lựa chọn và có tổng cộng 3711 kết quả bài thi môn Giải phẫu học, 2268 kết quả bài thi môn Sinh lý học đã được đưa vào phân tích. Độ khó, độ phân cách D 27%^(1,2), độ phân cách r_{pbis} , độ tin cậy theo Kuder Richardson⁽¹⁾ được xác định bởi phần mềm phân tích câu BasicStat⁽³⁾ của Đại học Y Dược TP. Hồ Chí Minh, các phần mềm thống kê và phân tích như Microsoft Excel, SPSS 20.

Kết quả: Đối với các đề thi môn Giải phẫu học, trung bình độ khó tương đối ổn định và có khuynh hướng dễ hơn qua các năm; trong khi trung bình độ phân cách r_{pbis} có khuynh hướng tăng dần, trung bình độ tin cậy ổn định ở mức tốt và có khuynh hướng tăng nhẹ; tỷ lệ những câu hỏi có độ phân cách kém hoặc nghi sai đáp án có khuynh hướng giảm dần. Đối với các đề thi môn Sinh lý học, trung bình độ khó tương đối ổn định qua các năm; trung bình độ phân cách r_{pbis} có khuynh hướng tăng nhẹ; trung bình độ tin cậy duy trì ở mức tốt và ổn định; tỷ lệ các câu hỏi khó có độ phân cách kém hoặc nghi sai đáp án cũng có khuynh hướng giảm dần.

Kết luận: Tất cả các bài thi tuyển sinh của hai môn Giải phẫu học và Sinh lý học đều cho thấy độ tin cậy cao. Qua các năm, độ khó của các bài thi này ổn định và có khuynh hướng giảm, trong khi đó khả năng phân biệt các nhóm thí sinh ở các trình độ cao và thấp tăng lên. Nghiên cứu cũng xác định một số lượng câu hỏi cần được cải thiện cũng có khuynh hướng giảm dần qua các năm. Phân tích câu hỏi thi cần được xây dựng thành quy trình nhằm cung cấp các thông tin cần thiết để nâng cao chất lượng đề thi tuyển sinh và xây dựng ngân hàng đề thi tuyển sinh sau đại học trong tương lai.

Từ khóa: độ khó, độ phân cách D 27%, độ phân cách r_{pbis} , lý thuyết trắc nghiệm cổ điển CTT

¹Bộ môn Tin học, Khoa Khoa học cơ bản, Đại học Y Dược TP. Hồ Chí Minh

²Phòng Đảm bảo chất lượng, Đại học Y Dược TP. Hồ Chí Minh

³Phòng Đào tạo Sau Đại học, Đại học Y Dược TP. Hồ Chí Minh

⁴Bộ môn Toán, Khoa Khoa học cơ bản, Đại học Y Dược TP. Hồ Chí Minh

⁵Trung tâm Y học Gia đình, Đại học Y Dược TP. Hồ Chí Minh

Tác giả liên lạc: ThS. Trần Thị Diệu

ĐT: 0918472230

Email: tranthidieu@ump.edu.vn

Tạp chí Y học Thành phố Hồ Chí Minh, 27(1):162-169. DOI: [10.32895/hcjm.m.2024.01.23](https://doi.org/10.32895/hcjm.m.2024.01.23)

ABSTRACT

ITEM ANALYSIS OF MULTIPLE CHOICE QUESTIONS OF ANATOMY AND PHYSIOLOGY ADMISSION TESTS FOR SPECIALIST LEVEL 1 IN UNIVERSITY OF MEDICINE AND PHARMACY AT HO CHI MINH CITY FROM 2018 TO 2023 USING CLASSICAL TEST THEORY

Tran Thi Dieu, Pham Duong Uyen Binh, Bui Tri Dung, Mai Phuong Thao, Nguyen Anh Vu, Vinh Son,
Pham Le An * Ho Chi Minh City Journal of Medicine * Vol. 27 - No. 1 - 2024: 162 – 169

Objectives: This study aimed to carry out a post-validation item analysis of multiple choice questions (MCQs) of Anatomy and Physiology in the admission test for Specialist level 1 programs based on item difficulty, item 27% discrimination (D 27%), item discrimination (r_{pbis}), and internal reliability in order to improve the quality of entrance exams.

Methods: A cross-sectional study was performed in the Department of Post-graduate training, University of Medicine and Pharmacy at Ho Chi Minh city. The results of 3711 Anatomy tests and 2268 Physiology tests, each consisting of 120 multiple-choice questions, collected from 2018 to 2023, have been analyzed. Item difficulty, discrimination index (D 27%), discrimination index (r_{pbis}), internal reliability according to Kuder Richardson were determined by BasicStat item analysis software and SPSS 20, Microsoft Excel.

Results: For the Anatomy MCQs, the average difficulty level was relatively stable and tended to slightly decrease. Whereas, the average discrimination index r_{pbis} showed a gradual increase, with the average reliability remaining at a high level and slightly increasing. The proportion of questions with poor discrimination or incorrect answers tended to decrease over time. For Physiology exam questions, the average difficulty level remained relatively stable over the years. The average r_{pbis} discrimination showed a slight increase. Average reliability kept a high and stable level; The rate of either difficult questions with poor separation or suspected incorrect answer choice also tended to decrease. The conducted item analysis revealed significant insights into the test's evolution. Despite items becoming easier over time, the test maintained high internal reliability, suggesting consistent precision in measuring the intended domain. Interestingly, items' increasing ease was accompanied with high discrimination indices, challenging the conventional belief that easy items lack of discrimination power.

Conclusions: All the Anatomy and Physiology tests demonstrated high internal reliability. Over the years, the difficulty of these tests decreased, but their discrimination power increased. A substantial number of MCQ items needed to be revised were reduced over times. Item analysis is necessary to pinpoint areas for improvement in the subsequent implementation of post-graduate admission tests.

Keywords: item difficulty, item 27% discrimination, r_{pbis} , reliability, Classical Test Theory

ĐẶT VẤN ĐỀ

Việc đánh giá kiến thức y khoa là một phần quan trọng của giáo dục y khoa, đặc biệt là trong các chương trình đào tạo y khoa sau đại học⁽¹⁾. Câu hỏi trắc nghiệm (MCQs) là một công cụ đo lường được sử dụng rộng rãi trong giáo dục y khoa do tính hiệu quả, độ tin cậy và khả năng đánh giá nhiều nội dung. Tuy nhiên, hiệu quả của MCQs trong việc đo lường chính xác kiến thức của người học còn phụ thuộc vào chất lượng của chính câu hỏi⁽²⁾.

Phân tích và đánh giá MCQs là một quá

trình thiết yếu để đảm bảo tính hợp lệ và độ tin cậy của MCQs được sử dụng trong kỳ thi tuyển sinh cho các chương trình đào tạo sau đại học y khoa⁽³⁾, cụ thể trong nghiên cứu này là của hai môn cơ sở Sinh lý học và Giải phẫu học dành cho đối tượng dự tuyển vào bậc đào tạo Chuyên khoa cấp I tại Đại học Y Dược TP. Hồ Chí Minh.

Lĩnh vực giáo dục y khoa xác định tầm quan trọng của việc đánh giá độ khó và độ phân cách của MCQs nhằm đưa ra quyết định đúng đắn về việc lựa chọn, sửa đổi hoặc loại bỏ các câu hỏi trong ngân hàng câu hỏi nói chung và đề thi nói

riêng. Độ khó của câu hỏi đề cập đến tỷ lệ thí sinh trả lời đúng một câu hỏi, trong khi độ phân cách của câu hỏi đánh giá khả năng phân biệt giữa thí sinh có năng lực cao và thí sinh có năng lực thấp đối với một câu hỏi MCQ. Việc đảm bảo rằng các MCQs có độ khó phù hợp và có khả năng phân biệt tốt năng lực của thí sinh sẽ góp phần nâng cao tính chính xác của đánh giá, nâng cao độ tin cậy và tính giá trị của đề thi tuyển sinh, góp phần tuyển chọn được người học có đủ năng lực tham gia vào quá trình đào tạo.

Với tầm quan trọng của các kỳ thi tuyển sinh đối với các chương trình đào tạo y khoa sau đại học, việc đánh giá kỹ lưỡng chất lượng của MCQs là điều cần thiết. Để đạt mục tiêu này, chúng tôi tiến hành phân tích câu MCQs được sử dụng trong kỳ thi tuyển sinh hai môn cơ sở là Sinh lý học và Giải phẫu học dành cho đối tượng dự tuyển bậc đào tạo Chuyên khoa cấp I.

Thông qua việc phân tích độ khó và độ phân cách của từng câu hỏi MCQ, nghiên cứu này mong muốn góp phần cải thiện việc đánh giá chất lượng, nâng cao tính giá trị, độ tin cậy của đề thi nói riêng và chất lượng tổng thể của các kỳ thi tuyển sinh đào tạo sau đại học về y khoa nói chung.

ĐỐI TƯỢNG-PHƯƠNG PHÁP NGHIÊN CỨU

Đối tượng nghiên cứu

Đây là một nghiên cứu cắt ngang được thực hiện trên tất cả các câu hỏi MCQs trong đề thi môn Giải phẫu học và Sinh lý học của các đợt thi tuyển sinh dành cho đối tượng Chuyên khoa cấp I từ năm 2018 đến năm 2023 tại Đại học Y Dược TP. Hồ Chí Minh. Mỗi đề thi có 120 câu hỏi trắc nghiệm. Nhóm nghiên cứu phân tích 3711 bài thi môn Giải phẫu học và 2268 bài thi môn Sinh lý học.

Câu hỏi trắc nghiệm được biên soạn theo Test Blue Print và được tiểu ban ra đề thi phê duyệt trước mỗi kỳ thi tuyển sinh đầu vào hàng năm. Mỗi câu hỏi trắc nghiệm chỉ có một lựa chọn đúng duy nhất.

Phương pháp nghiên cứu

Phân tích câu hỏi MCQs tập trung vào chỉ số độ khó (Dif), độ phân cách D 27%, độ phân cách r_{pbis} . Các câu hỏi có độ khó trong phạm vi từ 0,3 đến 0,7 và độ phân cách $r_{pbis} > 0,2$ thì được xem là câu hỏi phù hợp.

Phân tích câu và đề thi theo phương pháp lý thuyết trắc nghiệm cổ điển (CTT) và sử dụng phần mềm BasicStat⁽⁴⁾ tại Đại học Y Dược TP. Hồ Chí Minh, phần mềm thống kê SPSS 20 và Microsoft Excel. Đặc trưng câu hỏi dựa trên các yếu tố: độ khó p, độ phân cách D 27%, độ phân cách r_{pbis} .

Độ khó câu hỏi

Độ khó p bằng tỷ số phần trăm thí sinh làm đúng câu hỏi trên tổng số thí sinh tham gia làm câu hỏi đó. Độ khó p cho biết tỷ lệ thí sinh trả lời đúng câu hỏi nhiều hay ít. Giá trị p càng lớn thì câu hỏi càng dễ⁽⁷⁾.

$$\text{Độ khó } p = \frac{\sum \text{số thí sinh làm đúng}}{\sum \text{số thí sinh tham gia}}$$

Độ phân cách D 27%

Độ phân cách cho biết khả năng phân biệt thí sinh giỏi với thí sinh kém. Một đề thi trắc nghiệm muốn có độ phân cách tốt thì yêu cầu phải có nhiều câu hỏi với độ khó p ở mức trung bình. Khi đó điểm số thu được của nhóm thí sinh sẽ có phổ trải rộng⁽⁵⁾.

Phương pháp tính độ phân cách D 27%⁽⁶⁾ dựa vào tổng điểm thô của từng thí sinh. Nhóm cao gồm 27% thí sinh đạt điểm cao và nhóm thấp gồm 27% thí sinh đạt điểm thấp.

$D = (\text{Số đúng nhóm cao} - \text{Số đúng nhóm thấp}) / \text{Số thí sinh một nhóm}$.

hoặc,

$D = \text{Tỷ lệ trả lời đúng ở nhóm cao} - \text{Tỷ lệ trả lời đúng ở nhóm thấp}$.

Độ phân cách r_{pbis} (hệ số tương quan)

Độ phân cách r_{pbis} là một đại lượng đo lường mối liên hệ tuyến tính giữa 2 biến⁽⁵⁾. r_{pbis} là cách khác của phép tính độ phân cách D 27%. r_{pbis} của một câu hỏi là hệ số tương quan Pearson giữa các điểm của câu hỏi đó với tổng điểm của cả đề

thi xét trên mọi thí sinh làm đề trắc nghiệm; r_{pbis} có giá trị dương lớn (gần bằng 1) khi một thí sinh nào đó có điểm của câu hỏi cao và điểm của cả đề thi lại thấp và ngược lại.

Độ phân cách r_{pbis} có giá trị âm lớn (gần bằng -1) khi một thí sinh nào đó có điểm số của câu hỏi cao và điểm số của cả đề trắc nghiệm cũng cao và ngược lại. Thí sinh kém thì sẽ trả lời đúng nhiều, thí sinh giỏi thì sẽ trả lời sai nhiều.

Tuy nhiên trong trường hợp r_{pbis} có giá trị âm và p_{value} càng nhỏ thì câu hỏi càng có ý nghĩa thống kê.

Độ phân cách r_{pbis} bằng 0 nếu điểm của câu hỏi và điểm của cả đề trắc nghiệm không có mối liên hệ chặt chẽ và ổn định nào.

Trị số r_{pbis} của câu hỏi có thể chấp nhận được phải lớn hơn 0,2. Nhìn chung, câu hỏi có độ phân cách r_{pbis} tốt khi câu hỏi và cả đề trắc nghiệm đều đo lường cùng một thứ.

$$r_{pbis} = \frac{\bar{x}_2 - \bar{x}_1}{S_x} \sqrt{\frac{n_2 - n_1}{n^2}}$$

Công thức tính r_{pbis} với: $\bar{x}_1$ là điểm trung bình của số thí sinh trả lời đúng câu hỏi; $\bar{x}_2$ là điểm trung bình của số thí sinh trả lời sai câu hỏi; S_x là độ lệch chuẩn của điểm số (độ lệch chuẩn số câu đúng); n_1 : số thí sinh trả lời đúng câu hỏi; n_2 : số thí sinh trả lời sai câu hỏi; n : số thí sinh dự thi.

Độ tin cậy

Độ tin cậy của đề thi là đại lượng biểu thị

Bảng 1. Kết quả phân tích chung kết quả đề thi của hai môn cơ sở Giải phẫu học và Sinh lý học từ năm 2018 - 2023

Môn thi	Sinh lý học						Giải phẫu học					
Năm	2018	2019	2020	2021	2022	2023	2018	2019	2020	2021	2022	2023
Số thí sinh	549	692	914	781	992	1085	361	452	677	592	817	811
Độ tin cậy	0,9433	0,9423	0,9396	0,9520	0,9464	0,9438	0,9592	0,9580	0,9625	0,9624	0,9659	0,9678
Điểm trung bình												
Điểm	92,869	89,416	89,910	93,346	88,537	93,802	91,618	90,139	92,178	92,204	97,209	96,480
Tỷ lệ %	77,39%	74,51%	74,93%	77,79%	73,78%	78,17%	76,35%	75,12%	76,82%	76,84%	81,01%	80,40%
KTC 95%	91,532	88,173	88,855	92,201	87,468	92,840	89,615	88,317	90,642	90,601	95,853	95,081
	94,205	90,659	90,965	94,491	89,606	94,763	97,000	91,962	93,715	93,808	98,565	97,878
Độ khó trung bình												
Độ khó	0,774	0,745	0,749	0,778	0,738	0,782	0,763	0,751	0,768	0,768	0,810	0,804
KTC 95%	0,739	0,710	0,715	0,739	0,701	0,748	0,734	0,723	0,743	0,739	0,787	0,779
	0,809	0,780	0,784	0,817	0,775	0,815	0,793	0,779	0,793	0,798	0,833	0,829

mức độ chính xác của phép đo nhờ đề trắc nghiệm. Đề thi có độ tin cậy cao nghĩa là điểm số hầu như không thay đổi nếu thí sinh làm lại một đề tương đương. Công thức tính độ tin cậy theo phương pháp Kuder – Richardson⁽¹⁾ dựa trên ý tưởng xem mỗi câu trong đề trắc nghiệm là một đề trắc nghiệm tương đương, nghĩa là chúng có cùng điểm trung bình và cùng phương sai. Công thức K-R20⁽¹⁾ hơi khó áp dụng vì đòi hỏi phải biết độ khó p của từng câu hỏi.

Công thức tính độ tin cậy

L : số câu hỏi của đề thi; p_j : Tỷ lệ làm đúng câu hỏi thứ j ; $q_j=1-p_j$: Tỷ lệ làm sai đối với câu hỏi thứ j ; SD^2 : Phương sai của tổng điểm số mọi thí sinh đối với cả đề thi.

$$r = \frac{L}{L-1} \left(1 - \frac{\sum_{j=1}^L p_j q_j}{SD^2} \right)$$

Công thức tính phương sai

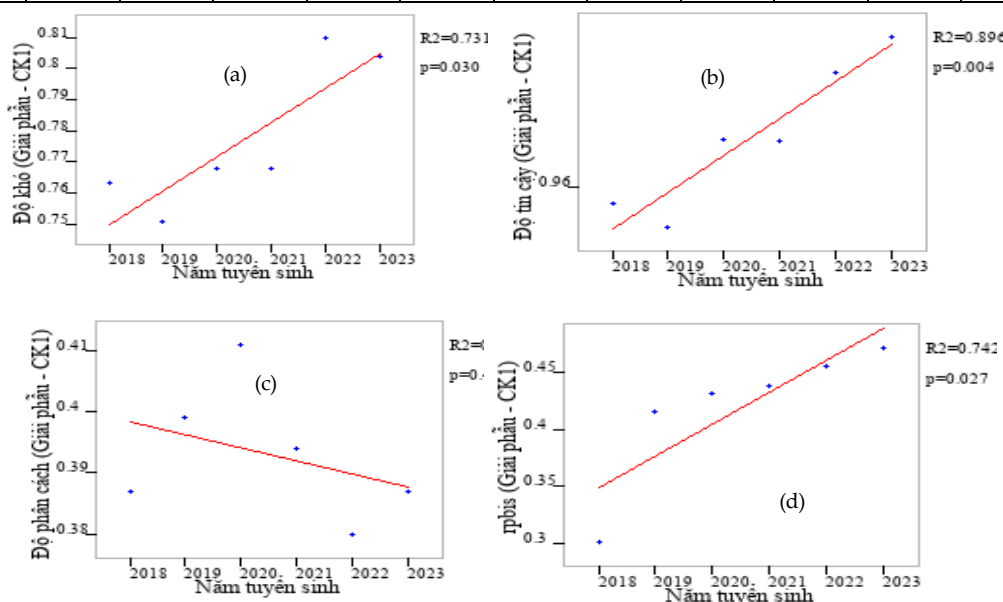
N : Số thí sinh; X_i : Điểm thí sinh i ; $\bar{X}$: Điểm trung bình.

$$SD^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}$$

KẾT QUẢ

Số lượng thí sinh thi môn thi Giải phẫu học và Sinh lý học tăng hơn gấp đôi từ năm 2018 đến năm 2023 (Bảng 1). Điểm trung bình hai môn này ở mức cao, dao động từ 88,53 đến 93,8 ở môn Sinh lý học và từ 90,13 đến 97,2 ở môn Giải phẫu học.

Môn thi	Sinh lý học						Giải phẫu học					
Độ phân cách D 27% trung bình												
D 27%	0,313	0,331	0,321	0,310	0,340	0,311	0,387	0,399	0,411	0,394	0,380	0,387
KTC 95%	0,279	0,301	0,293	0,279	0,310	0,281	0,358	0,370	0,384	0,365	0,353	0,361
	0,348	0,362	0,349	0,342	0,369	0,342	0,416	0,427	0,439	0,423	0,407	0,412
Độ phân cách r_{pbis} trung bình												
r_{pbis}	0,300	0,359	0,349	0,383	0,376	0,367	0,302	0,415	0,432	0,438	0,455	0,472
KTC 95%	0,262	0,334	0,325	0,354	0,350	0,341	0,262	0,392	0,414	0,418	0,434	0,454
	0,338	0,385	0,374	0,413	0,402	0,392	0,343	0,438	0,450	0,458	0,476	0,490



Hình 1. Tương quan giữa (a) độ khó, (b) độ tin cậy, (c) độ phân cách D 27%, (d) r_{pbis} đối với năm tuyển sinh môn thi Giải phẫu học

Đối với các đề thi tuyển Chuyên khoa cấp I môn thi Giải phẫu học, Bảng 1 cho thấy độ khó trung bình tương đối ổn định, có giá trị dao động từ 0,751 (75,1%) đến 0,81 (81%) và có khuynh hướng giảm nhẹ ($r=0,86$, $p=0,030$) với hệ số góc $\alpha=0,011$. Độ tin cậy có giá trị dao động từ 0,95 đến 0,96 và cũng có khuynh hướng tăng nhẹ ($r=0,947$, $p=0,004$) với hệ số góc $\alpha=0,002$.

Hình 1 (a) và (d) cho thấy mặc dù độ khó giảm qua các năm nhưng độ phân cách r_{pbis} vẫn tăng ($r=0,861$, $p=0,0137$) với hệ số góc $\alpha=0,03$; trong khi đó (a) và (c) cho thấy độ khó giảm và độ phân cách D 27% cũng giảm ($r=-0,36$) khác biệt không có ý nghĩa thống kê ($p=0,483$).

Như vậy so sánh giữa (c) và (d), sử dụng độ phân cách r_{pbis} chính xác hơn so với độ phân cách D 27% đặc biệt đối với những đề thi được đánh giá là quá dễ. Điều này nghĩa là khi đề càng dễ thì càng nên sử dụng độ phân cách r_{pbis} . Nhấn

manh, độ phân cách D 27% chỉ dành cho nhóm lớp dạy của môn học và khi sử dụng độ phân cách r_{pbis} thì có thể giúp gia tăng số câu hỏi tốt trong ngân hàng đề.

Thêm một bằng chứng quan sát (c) và (b), liệu rằng có hay không “độ phân cách D 27% giảm nhưng độ tin cậy lại tăng?” Giả thiết truyền thống khẳng định khi độ phân cách tăng thì độ tin cậy sẽ tăng và ngược lại. Do đó, có thể khẳng định việc sử dụng độ phân cách r_{pbis} là hợp lý, đặc biệt đối với những đề thi dễ.

Bảng 2 cho thấy tỷ lệ câu có độ khó $\leq 0,3$ và độ phân cách $r_{pbis} < 0,2$ có khuynh hướng giảm dần ($r=-0,861$, $p=0,0139$) từ 4 câu (3,33%) đến 0 câu (0%).

Tỷ lệ câu dễ (độ khó $\geq 0,8$) có độ phân cách r_{pbis} tốt ($\geq 0,4$) tăng dần qua các năm ($r=0,976$, $p=0,000886$) với hệ số góc $\alpha=6,999$,

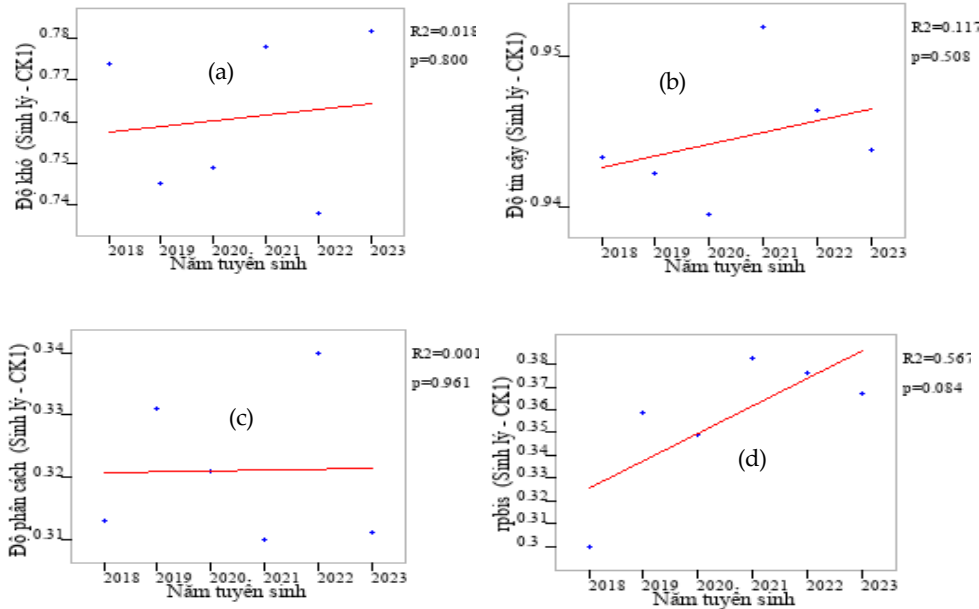
cho thấy mặc dù đề thi dễ dần nhưng tỷ lệ câu có độ phân cách tốt lại tăng lên. Bên cạnh đó, đề thi có độ tin cậy cao hơn do tỷ lệ các câu có độ phân cách r_{pbis} chấp nhận được có khuynh hướng tăng dần qua các năm từ 31% đến 62%. Qua các năm nghiên cứu đề thi Giải phẫu học, các câu có độ phân cách r_{pbis} chưa tốt ($<0,2$) tập trung chủ yếu các câu

khó; các câu có độ phân cách r_{pbis} tốt $\geq 0,4$ thì không có câu khó nhưng đáng chú ý hơn là câu dễ có khuynh hướng tăng dần từ tỷ lệ từ 21,67% đến 54,17%.

Theo quan niệm truyền thống, đề thi càng dễ thì độ phân cách càng kém. Tuy nhiên, từ kết trên có thể thấy đề thi dễ hơn qua các năm nhưng độ phân cách có khuynh hướng tăng dần.

Bảng 2. Phân bố câu hỏi theo độ khó và độ phân cách r_{pbis} của môn thi Giải phẫu học

r_{pbis}	<0,2				0,2 ~ 0,4				≥ 0,4				Tổng
Độ khó	≥ 0,8		≤ 0,3		≥ 0,8		≤ 0,3		≥ 0,8		≤ 0,3		
Năm	Số câu	Tỷ lệ (%)	Số câu	Tỷ lệ (%)	Số câu	Tỷ lệ (%)	Số câu	Tỷ lệ (%)	Số câu	Tỷ lệ (%)	Số câu	Tỷ lệ (%)	
2018	14	11,67	4	3,33	12	10	0	0	26	21,67	0	0	56
2019	0	0	2	1,67	20	16,67	0	0	29	24,17	0	0	51
2020	0	0	1	0,83	21	17,5	0	0	33	27,5	0	0	55
2021	0	0	2	1,67	16	13,33	0	0	45	37,5	0	0	63
2022	0	0	1	0,83	16	13,33	0	0	58	48,33	0	0	75
2023	1	0,83	0	0	10	8,33	0	0	65	54,17	0	0	76



Hình 2. Tương quan giữa (a) độ khó, (b) độ tin cậy, (c) độ phân cách D 27%, (d) r_{pbis} đối với năm tuyển sinh môn thi Sinh lý học

Đối với các đề thi môn Sinh lý học, Bảng 1 cho thấy độ khó trung bình tương đối ổn định, có giá trị dao động từ 0,732 (73,2%) đến 0,738 (73,8%) và có khuynh hướng giảm rất ít, không đáng kể ($r=0,134$, $p=0,799$) với hệ số góc $a=0,001$. Độ tin cậy có giá trị dao động từ 0,94 đến 0,95 và cũng có khuynh hướng tăng nhẹ ($r=0,341$, $p=0,508$) với hệ số góc $a=0,0008$.

Hình 2 (a) và (d) cho thấy mặc dù độ khó giảm qua các năm nhưng độ phân cách r_{pbis} vẫn tăng ($r=0,753$, $p=0,042$) với hệ số góc $a=0,012$; trong khi đó (a) và (c) cho thấy độ khó giảm và độ phân cách D 27% tăng không đáng kể ($r=0,0026$, $p=0,48$) với hệ số góc $a=0,00017$.

Kết quả so sánh giữa (c) và (d) cho thấy việc sử dụng độ phân cách r_{pbis} chính xác hơn so với

độ phân cách D 27% đặc biệt đối với những đề thi dễ và quá dễ.

Bảng 3. Bảng phân bố câu hỏi theo độ khó và độ phân cách r_{pbis} của môn thi Sinh lý học.

r_{pbis}	< 0,2				0,2 ~ 0,4				$\geq 0,4$				Tổng
Độ khó	$\geq 0,8$		$\leq 0,3$		$\geq 0,8$		$\leq 0,3$		$\geq 0,8$		$\leq 0,3$		
Năm	Số câu	Tỷ lệ (%)	Số câu	Tỷ lệ (%)	Số câu	Tỷ lệ (%)	Số câu	Tỷ lệ (%)	Số câu	Tỷ lệ (%)	Số câu	Tỷ lệ (%)	
2018	11	9,17	7	5,83	29	24,17	0	0	29	24,17	0	0	76
2019	6	5	6	5	33	27,5	0	0	23	19,17	0	0	68
2020	1	0,83	0	0	39	32,5	1	0,83	20	16,67	0	0	61
2021	6	5	5	4,17	33	27,5	1	0,83	34	28,33	0	0	79
2022	4	3,33	7	5,83	28	23,33	0	0	33	27,5	0	0	72
2023	6	5	5	4,17	35	29,17	0	0	34	28,33	0	0	80

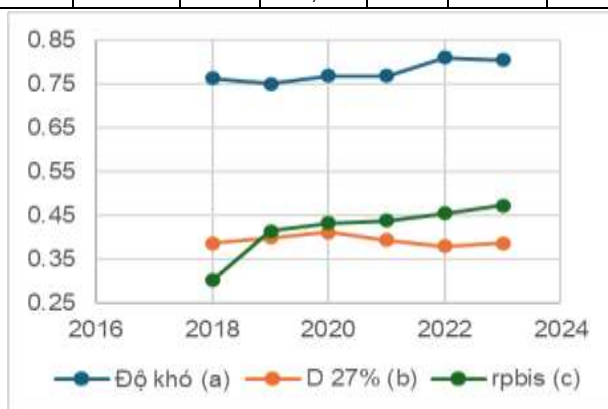
Bảng 3 cho thấy tỷ lệ câu có độ khó $\leq 0,3$ và độ phân cách $r_{pbis} < 0,2$ dao động trong khoảng từ 7 câu (5,38%) đến 0 câu (0%). Tỷ lệ câu dễ (độ khó $\geq 0,8$) có độ phân cách r_{pbis} tốt ($\geq 0,4$) tăng dần qua các năm ($r=0,61$, $p=0,199$) với hệ số góc $a=1,64$.

Qua các năm nghiên cứu đề thi môn Sinh lý học, chúng tôi ghi nhận, đề thi có độ tin cậy cao do tỷ lệ các câu có độ phân cách r_{pbis} chấp nhận được chiếm tỷ lệ từ 48% đến 56%. Các câu có độ phân cách r_{pbis} chưa tốt ($< 0,2$) tập trung cả câu khó và câu dễ. Các câu có độ phân cách r_{pbis} tốt ($\geq 0,4$) không có câu khó, đáng chú ý là câu dễ chiếm tỷ lệ từ 16% đến 28% trong các câu có độ phân cách tốt và có khuynh hướng tăng dần qua các năm.

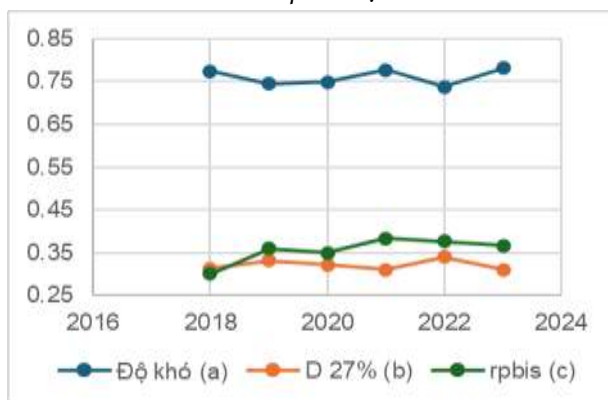
Độ phân cách D 27% và r_{pbis}

Hình 3 (b) độ phân cách D 27% trung bình dao động trong khoảng giá trị từ 0,38 đến 0,411 và có khuynh hướng giảm dần ($r=-0,36$, $p=0,483$) với hệ số góc $a=-0,002$. Trong khi đó, (c) độ phân cách r_{pbis} trung bình có giá trị dao động từ 0,302 đến 0,472 và có khuynh hướng tăng dần ($r=0,861$, $p=0,028$) với hệ số góc $a=0,03$.

Hình 4 (b) cho thấy độ phân cách D 27% trung bình dao động trong khoảng giá trị từ 0,31 đến 0,34 và có khuynh hướng tăng rất ít, không đáng kể ($r=0,026$, $p=0,961$) với hệ số góc $a=0,00017$. Trong khi đó, r_{pbis} trung bình có giá trị dao động từ 0,3 đến 0,382 và có khuynh hướng tăng dần ($r=0,753$, $p=0,084$) với hệ số góc $a=0,012$.



Hình 3. So sánh trung bình r_{pbis} và D 27% của môn Giải phẫu học



Hình 4. So sánh trung bình r_{pbis} và D 27% của môn Sinh lý học

BÀN LUẬN

Đề thi có độ phân cách r_{pbis} cao ($\geq 0,4$) cho biết các câu hỏi có khả năng phân biệt tốt năng lực của thí sinh, trình độ của thí sinh chênh lệch nhau nhiều, đề thi có độ tin cậy cao, và cuối cùng là đề thi có độ giá trị cao.

Qua các năm, độ tin cậy của các đề thi vẫn luôn ổn định ở mức cao và dao động từ 0,9396

đến 0,9878. Yếu tố quan trọng giúp cho đề thi có độ tin cậy cao và ổn định là độ phân cách của đề thi tốt. Độ phân cách r_{pbis} của các đề thi qua các năm có khuynh hướng tăng dần và luôn ở mức tốt. Ngoài ra, số câu hỏi cần cải thiện ($r_{pbis} < 0,2$ và độ khó câu hỏi $\leq 0,3$) cũng có khuynh hướng giảm dần. Điều này chứng tỏ rằng đề thi càng ngày càng có chất lượng hơn, góp phần tăng thêm khả năng đánh giá năng lực của thí sinh hơn.

Độ tin cậy của đề thi cao ($r > 0,8$) có thể do số câu hỏi nhiều (120 câu); độ khó đề thi phù hợp; độ phân cách r_{pbis} cao. Tất cả các yếu tố này góp phần làm cho đề thi có độ giá trị cao.

Mặc dù đề thi dễ dần nhưng vẫn có độ phân cách r_{pbis} cao. Đối với đề thi dễ, khuyến cáo chỉ dùng độ phân cách D 27% để lựa chọn câu; tuy nhiên chỉ giới hạn độ phân cách D 27% dành cho phạm vi lớp học, còn riêng đối với độ phân cách r_{pbis} thì được sử dụng cho mọi cấp độ lựa chọn câu như khoa, trường, quốc gia do sử dụng toàn bộ số liệu chứ không phải chỉ 27% như độ phân cách D.

Từ kết quả phân tích ở Bảng 2 và Bảng 3 cho thấy đề thi dễ dần nhưng độ phân cách r_{pbis} có khuynh hướng tăng dần. Trong khi đó, theo quan niệm truyền thống, đề thi càng dễ thì độ phân cách càng kém⁽⁷⁾. Như vậy, chúng tôi đặt ra câu hỏi vì sao các đề thi có độ khó giảm dần nhưng độ phân cách r_{pbis} có khuynh hướng tăng dần?

Từ kết quả nghiên cứu này, chúng tôi sẽ tiếp tục nghiên cứu những vấn đề liên quan:

Liệu rằng còn có những yếu tố nào khác ảnh hưởng đến kết quả này không? Ví dụ như mối liên hệ giữa độ khó, độ phân cách r_{pbis} của đề thi đối với các yếu tố liên quan khác như năm tốt nghiệp, xếp loại tốt nghiệp, trường tốt nghiệp đại học, v.v... của thí sinh dự thi.

Phân tích mỗi nhữ để nhận biết các sai lầm phổ biến, mỗi nhữ kém, nhận biết câu hỏi tiêu chí/câu hỏi tiêu chuẩn,... giúp cải thiện và nâng cao chất lượng câu hỏi thi ngày càng tốt hơn và hướng đến việc xây dựng ngân hàng đề thi tuyển sinh sau đại học trong tương lai.

KẾT LUẬN

Tất cả các đề thi Giải phẫu học và Sinh lý học dành cho thí sinh dự tuyển chương trình đào tạo Chuyên khoa cấp I đều cho thấy độ tin cậy cao. Qua các năm, độ khó của các đề thi này có khuynh hướng giảm, trong khi khả năng phân biệt các nhóm thí sinh ở các trình độ cao và thấp tăng lên. Nghiên cứu cũng chỉ ra số lượng câu hỏi cần được cải thiện cũng giảm dần qua các năm. Phân tích câu MCQs thường quy tại các bộ môn cung cấp các thông tin cần thiết để cải thiện các bài thi tuyển sinh sau đại học trong tương lai.

TÀI LIỆU THAM KHẢO

1. Kheyami D, Jaradat A, Al-Shibani T, Ali FA (2018). Item Analysis of Multiple Choice Questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos Univ Med J*. 18(1):e68-e74. 10.18295/squmj.2018.18.01.011.
2. Robert LE, David AF (1991). *Essentials of educational measurement*, University of Iowa.
3. Özkan YÖ, Güvendir MA (2021). Differential Item Functioning Analysis of a High Stake Test in Terms of Statistical Regions of Turkey. *Journal of Pedagogical Research*. 5(3):122-134.
4. Đại học Y Dược Thành phố Hồ Chí Minh (2023). Phân tích câu hỏi và đề thi trắc nghiệm. URL: <https://basicstat.net/mcq2023/>.
5. Lâm Quang Thiệp (2011), *Đo lường trong Giáo dục*, NXB Đại học quốc gia Hà Nội.
6. Dương Thiệu Tống (2005). *Thống kê ứng dụng trong nghiên cứu khoa học giáo dục*. NXB Khoa học Xã hội.
7. De C, José F (2014). Classical test theory and item response theory: Two understandings of one high-stakes performance exam1. *Colombian Applied Linguistics Journal*, 16(2):167-184.

Ngày nhận bài: 17/04/2024

Ngày chấp nhận đăng bài: 10/06/2024

Ngày đăng bài: 11/06/2024