ĐÁNH GIÁ ĐỘ KHÓ, ĐỘ PHÂN BIỆT MỘT SỐ ĐỀ THI SỬ DỤNG TẠI TRƯỜNG ĐẠI HỌC Y KHOA VINH

Lê Thị Mai Hoa

Trường Đại học Y khoa Vinh

Đánh giá chất lượng đề thi trắc nghiệm khách quan sau kỳ thi giữ vai trò then chốt trong việc bảo đảm tính giá trị và độ tin cậy của công cụ đánh giá kết quả học tập. Nghiên cứu được tiến hành nhằm phân tích độ khó và độ phân biệt của các đề thi kết thúc học phần Phụ sản 1, Dinh dưỡng tiết chế, Chăm sóc sức khỏe trẻ em, Dược lý 1 và Hóa sinh 1 tại trường Đại học Y khoa Vinh. Thiết kế nghiên cứu mô tả cắt ngang, phân tích hồi cứu kết quả thi lý thuyết trắc nghiệm cổ điển bằng phần mềm Quản lý đào tạo Education (E). Kết quả phân tích cho thấy chất lượng đề thi có sự khác biệt đáng kể giữa các học phần. Ở học phần Phụ sản 1, khoảng 50% câu hỏi có độ khó trung bình (0,2 - 0,8), 35% là câu hỏi dễ và 15% là câu hỏi khó; về độ phân biệt, 60% câu hỏi đạt yêu cầu (D ≥ 0,2); tuy nhiên, vẫn còn 40% câu hỏi có độ phân biệt thấp cần được chỉnh sửa. Đối với học phần Dinh dưỡng tiết chế, tỷ lệ câu hỏi có độ khó trung bình rất cao (79,9%), phản ánh đề thi tập trung chủ yếu ở mức trung bình và khả năng phân hóa chưa rõ ràng. Học phần Chăm sóc sức khỏe trẻ em có 67,3% câu hỏi ở mức trung bình, trong khi tỷ lệ câu hỏi dễ tương đối thấp (7,6%). Ở học phần Dược lý 1, tỷ lệ câu hỏi trung bình đạt 57,4%, phù hợp với yêu cầu của một đề thi chuẩn. Ngược lại, học phần Hóa sinh 1 có tới 83,4% câu hỏi ở mức trung bình, cao hơn ngưỡng khuyến nghị. Nhìn chung, phân tích đề thi sau kỳ thi giúp nhận diện chính xác các câu hỏi chưa đạt chuẩn, làm cơ sở cho việc sàng lọc, điều chỉnh và chuẩn hóa ngân hàng câu hỏi, qua đó góp phần nâng cao chất lượng đánh giá trong đào tạo y khoa.

Từ khóa: độ khó, độ phân biệt, trắc nghiệm khách quan, đánh giá đề thi.

I. ĐẶT VẤN ĐỀ

Trong giáo dục y khoa hiện đại, trắc nghiệm khách quan (TNKQ) đóng vai trò quan trọng trong việc đánh giá kiến thức của người học nhờ khả năng bao phủ nội dung rộng và tính khách quan cao [1]. Tuy nhiên, để đảm bảo tính chính xác của công cụ đánh giá kiến thức, các câu hỏi thi cần được đánh giá định lượng thông qua các tham số đặc trưng, quan trọng nhất là độ khó (Difficulty index - P) và độ phân biệt (Discrimination index - D) [2].

Tại trường Đại học Y khoa Vinh, việc ứng dụng phần mềm Quản lý đào tạo Education (E) đã cho phép trích xuất các dữ liệu này một cách hệ thống [3]. Mặc dù quy trình biên soạn đề thi đã được thực hiện chặt chẽ, việc đánh giá lại thực tế chất lượng câu hỏi dựa trên dữ liệu thi thật vẫn là yêu cầu cấp thiết. Chúng tôi thực hiện nghiên cứu này nhằm mục tiêu đánh giá độ khó, độ phân biệt một số đề thi sử dụng tại trường Đại học Y khoa Vinh.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

1. Đối tượng nghiên cứu

Nghiên cứu tiến hành phân tích 5 đề thi kết thúc học phần TNKQ (4 lựa chọn), các câu hỏi và kết quả làm bài của sinh viên được sử dụng trong năm 2024-2025, bao gồm:

1. Học phần Phụ sản 1 (ngành Y khoa, lớp 20YA, mã đề 0500000547, 100 câu).

2. Học phần Dinh dưỡng tiết chế (ngành Y học dự phòng, lớp YHDP8, mã đề 0503000888).

3. Học phần Chăm sóc sức khỏe trẻ em (ngành Điều dưỡng, lớp Đ12A, mã đề 0503000854).

4. Học phần Dược lý 1 (ngành Dược học, lớp D4A, mã đề 0503000831).

5. Học phần Hóa sinh 1 (ngành Kỹ thuật xét nghiệm y học, lớp Đ8XN, mã đề 0503000979).

2. Phương pháp nghiên cứu

Thiết kế nghiên cứu: mô tả cắt ngang.

Tác giả chính: *Lê Thị Mai Hoa*
Email: maihoa@vmu.edu.vn

Xử lý số liệu: sử dụng phần mềm Quản lý đào tạo Education (E).

Tiêu chuẩn đánh giá:

* Độ khó (P): độ khó của câu hỏi có thể chấp nhận được trong khoảng 0,2 ≤ P ≤ 0,8; Câu hỏi có độ khó > 0,8 là quá dễ; Câu hỏi có độ khó < 0,2 là quá khó. Mức khuyến nghị cho đề thi tốt là đa số câu hỏi nằm trong khoảng 0,2 - 0,8 [4].

* Độ phân biệt (D): độ phân biệt D ≥ 0,4: câu hỏi có độ phân biệt rất tốt; 0,30 ≤ D ≤ 0,39: câu hỏi có độ phân biệt tốt; 0,20 ≤ D ≤ 0,29: câu hỏi có độ phân biệt có thể chấp nhận được; D < 0,2: câu hỏi có độ phân biệt chưa đạt yêu cầu, phải loại bỏ hoặc thay đổi để nâng cao độ phân cách [4].

* Công thức Kuder-Richardson-20 như sau để tính độ tin cậy của một đề trắc nghiệm:

$$r = \frac{k}{k-1}(1 - \frac{\sum_{i=1}^{k} p_i q_i}{\sigma^2}$$

Trong đó:

k: số câu hỏi của đề trắc nghiệm;

pi: tỷ lệ trả lời đúng đối với câu hỏi thứ i;

qi = (1-pi): tỷ lệ trả lời sai đối với câu hỏi thứ i;

σ2: phương sai của tổng điểm mọi thí sinh đối với cả đề trắc nghiệm.

III. KẾT QUẢ NGHIÊN CỨU

1. Đặc điểm chung của đề thi

Bảng 1. Tổng hợp thông tin chung của các đề thi

Tên học phần	Số câu hỏi	Số sinh viên dự thi	Độ khó trung bình (P)	Độ phân biệt trung bình (D)
Phụ sản 1	100	50	0,58	0,23
Dinh dưỡng tiết chế	80	14	0,50	0,25
CS sức khỏe trẻ em	80	57	0,41	0,15
Dược lý 1	80	39	0,66	0,29
Hóa sinh 1	60	25	0,39	0,14

Các đề thi có phân bố lệch trái: một số câu hỏi có độ khó rất thấp (câu quá khó).

2. Nhận xét về độ khó (P)

Bảng 2. Phân bố tỷ lệ câu hỏi theo mức độ khó (%)

Học phần	Khó (P < 0,2)	Trung bình (0,2 - 0,8)	Dễ (P > 0,8)
Phụ sản 1	15	50	35
Dinh dưỡng tiết chế	8,8	79,9	11,3
CS sức khỏe trẻ em	25,1	67,3	7,6
Dược lý 1	8,8	57,4	33,8
Hóa sinh 1	16,6	83,4	0

Đối với học phần Phụ sản 1, kết quả phân tích trên 100 câu hỏi cho thấy tỷ lệ câu hỏi ở mức độ trung bình chiếm 50%, đạt yêu cầu của một đề thi tiêu chuẩn. Tuy nhiên, tỷ lệ câu hỏi dễ còn khá cao (35%), trong khi tỷ lệ câu hỏi khó chiếm 15%. Điều này cho thấy đề thi có xu hướng hơi dễ so với năng lực của nhóm sinh viên dự thi (Lớp 20YA).

Đối với học phần Dinh dưỡng tiết chế, kết quả phân tích trên 80 câu hỏi cho thấy tỷ lệ câu hỏi ở mức độ trung bình chiếm 79,9%; đây là tỷ lệ cao, chứng tỏ đề thi đang ở mức trung bình chưa phân hóa rõ.

Đối với học phần Chăm sóc sức khỏe trẻ em, kết quả phân tích trên 80 câu hỏi cho thấy tỷ lệ câu hỏi ở mức độ trung bình chiếm 67,3 %; hơi cao so với yêu cầu, trong đó câu hỏi dễ còn hơi ít (7,6%). Điều này cho thấy đề thi có xu hướng hơi khó so với năng lực của nhóm sinh viên dự thi.

Đối với học phần Dược lý 1, kết quả phân tích trên 80 câu hỏi cho thấy tỷ lệ câu hỏi ở mức độ trung bình chiếm 57,4%, đạt yêu cầu của một đề thi tiêu chuẩn. Tuy nhiên, tỷ lệ câu hỏi dễ còn khá cao (33,8 %). Điều này cho thấy đề thi có xu hướng hơi dễ so với năng lực của nhóm sinh viên dự thi.

Đối với học phần Hóa sinh 1, kết quả phân tích trên 60 câu hỏi cho thấy tỷ lệ câu hỏi ở mức độ trung bình chiếm 83,4 % cao so với yêu cầu, trong đó câu hỏi dễ không có. Điều này cho thấy đề thi hơi khó so với năng lực của nhóm sinh viên dự thi.

3. Nhận xét về độ phân biệt (D)

Bảng 3. Phân bố tỷ lệ câu hỏi theo độ phân biệt (%)

Học phần	Tốt & Rất tốt (D ≥ 0,3)	Chấp nhận được (0,2 ≤ D < 0,3)	Kém (D < 0,2)
Phụ sản 1	30	30	40
Dinh dưỡng tiết chế	38	26	36
Chăm sóc sức khỏe trẻ em	22,6	24,8	52,6
Dược lý 1	46,4	12,3	41,3
Hóa sinh 1	18,3	21,7	60

Tại học phần Phụ sản 1, tổng tỷ lệ câu hỏi có độ phân biệt đạt yêu cầu (từ mức chấp nhận được trở lên) là 60%. Đáng chú ý, có tới 40% câu hỏi có độ phân biệt kém (D<0,2). Phân tích chi tiết cho thấy nhiều câu hỏi trong nhóm này là các câu quá dễ (ví dụ câu 29, 31, 41, 46 có P>0,9 và D<0,1), khiến cho hầu hết sinh viên đều làm đúng, do đó không phân loại được sinh viên giỏi và kém. Ngoài ra, một số câu hỏi có độ phân biệt âm (ví dụ câu 15, 69, 76), gợi ý rằng có thể có vấn đề về nội dung câu hỏi hoặc đáp án gây nhiễu chưa hợp lý. Học phần Dinh dưỡng tiết chế, tổng tỷ lệ câu hỏi có độ phân biệt đạt yêu cầu (từ mức chấp nhận được trở lên) là 64%. Học phần Chăm sóc sức khỏe trẻ em tổng tỷ lệ câu hỏi có độ phân biệt đạt yêu cầu (từ mức chấp nhận được trở lên) là 47,4%. Học phần Dược lý 1 tổng tỷ lệ câu hỏi có độ phân biệt đạt yêu cầu (từ mức chấp nhận được trở lên) là 58,7%. Học phần Hóa sinh 1 tổng tỷ lệ câu hỏi có độ phân biệt đạt yêu cầu (từ mức chấp nhận được trở lên) là 40%.

IV. BÀN LUẬN

1. Về độ khó

Độ khó trung bình được chấp nhận của một đề thi trắc nghiệm (4 lựa chọn) nên nằm trong khoảng 0,2 đến 0,8 [1]. Kết quả nghiên cứu cho thấy học phần Phụ sản 1 cơ bản đảm bảo được ma trận này, với 50% câu hỏi ở mức trung bình. Tuy nhiên, tỷ lệ 35% câu hỏi dễ là khá cao, có thể làm giảm khả năng phân loại ở nhóm sinh viên khá - giỏi. Các câu hỏi quá dễ (P>0,8) thường rơi vào mức độ "Nhớ", cần được xem xét để nâng cao mức độ nhận thức lên "Hiểu" hoặc "Vận dụng".

2. Về độ phân biệt

Độ phân biệt là chỉ số quan trọng nhất, phản ánh chất lượng câu hỏi. Theo Ebel và Frisbie (1991), các câu hỏi có D < 0,2 cần được loại bỏ hoặc chỉnh sửa kỹ lưỡng [5]. Kết quả phân tích học phần Phụ sản 1 cho thấy, tỷ lệ câu hỏi có độ phân biệt kém còn cao (40%). Tỷ lệ câu hỏi trong học phần Dinh dưỡng tiết chế có độ phân biệt kém là 40%. Học phần Chăm sóc sức khỏe trẻ em, tỷ lệ câu hỏi có độ phân biệt kém là 52,6%. Học phần Dược lý 1 có tỷ lệ câu hỏi độ phân biệt kém là 41,3%. Còn học phần Hóa sinh 1 tỷ lệ đó là 60%.

Nguyên nhân của các câu hỏi có độ phân biệt thấp thường do: (1) câu hỏi quá dễ hoặc quá khó (hiệu ứng sàn hoặc trần); (2) câu hỏi có nội dung không rõ ràng, gây hiểu nhầm; (3) phương án nhiễu không hoạt động hiệu quả (không thu hút được sinh viên kém); (4) lỗi kỹ thuật trong đáp án (sai đáp án trên phần mềm); (5) các câu hỏi có độ phân biệt âm (sinh viên giỏi làm sai nhiều hơn

sinh viên kém) là những câu hỏi "báo động đỏ", cần được kiểm tra ngay lập tức về đáp án và nội dung chuyên môn.

V. KẾT LUẬN

Qua phân tích đề thi học phần Phụ sản 1 và các học phần liên quan, chúng tôi nhận thấy quy trình đánh giá định lượng bằng phần mềm Education (E) đã cung cấp bức tranh rõ ràng về chất lượng ngân hàng câu hỏi. Các bộ môn khác cần rà soát lại các câu hỏi thuộc nhóm có độ phân biệt kém (D < 0,2) và các câu hỏi quá dễ. Cần đa dạng hóa các câu hỏi mức độ vận dụng để nâng cao độ phân biệt của đề thi trong các năm học tiếp theo.

TÀI LIỆU THAM KHẢO

[1] **Lâm Quang Thiệp** (2011), *Đo lường trong giáo dục: Lý thuyết và Ứng dụng*, NXB Đại học Quốc gia Hà Nội.

[2] **Bộ Giáo dục và Đào tạo**, *Quy chế thi trung học phổ thông quốc gia.*

[3] **Hướng dẫn sử dụng phần mềm quản lý đào tạo E tại trường Đại học Y khoa Vinh**, *Tài liệu chuyển giao công nghệ phần mềm.*

[4] **Dương Thiệu Tống** (2000), *Thống kê ứng dụng trong nghiên cứu khoa học giáo dục*, NXB Đại học Quốc gia Hà Nội.

[5] *Ebel R.L., Frisbie D.A*. (1991), *Essentials of Educational Measurement*, Prentice Hall.

SUMMARY

TO ASSESS THE DIFFICULTY INDEX AND DISCRIMINATION INDEX OF SEVERAL END-OF-COURSE EXAMINATIONS USED AT VINH UNIVERSITY OF MEDICINE

Post-examination evaluation of multiple-choice question (MCQ) tests plays a pivotal role in ensuring the validity and reliability of assessment tools used to measure learning outcomes. This study was conducted to analyze the difficulty and discrimination indices of end-of-course examinations in Obstetrics 1, Nutrition and Dietetics, Child Health Care, Pharmacology 1, and Biochemistry 1 at Vinh Medical University. A cross-sectional descriptive design with retrospective analysis was applied, using Classical Test Theory to examine examination results processed through the Education (E) training management software. The analysis revealed significant differences in test quality among the courses. In Obstetrics 1, approximately 50% of the items demonstrated moderate difficulty levels (0.2–0.8), while 35% were easy and 15% were difficult; in terms of discrimination, 60% of the items met acceptable criteria (D ≥ 0.2), whereas 40% showed low discrimination and required revision. For the Nutrition and Dietetics course, the proportion of moderately difficult items was very high (79.9%), indicating that the test was heavily concentrated at an average difficulty level with limited discriminatory power. The Child Health Care course exhibited 67.3% moderately difficult items, while the proportion of easy questions was relatively low (7.6%). In Pharmacology 1, moderately difficult items accounted for 57.4%, consistent with the requirements of a standard examination. In contrast, Biochemistry 1 showed an excessively high proportion of moderately difficult items (83.4%), exceeding recommended thresholds. Overall, post-examination item analysis enables accurate identification of substandard questions and provides a foundation for item screening, revision, and standardization of the question bank, thereby contributing to improved assessment quality in medical education.

Keywords: Difficulty index, discrimination index, multiple-choice questions, examination evaluation.