

## ĐÁNH GIÁ CHẤT LƯỢNG CÂU HỎI VÀ ĐỀ THI TRẮC NGHIỆM KHÁCH QUAN CỦA CÁC HỌC PHẦN NỘI KHOA, NGOẠI KHOA TẠI HỌC VIỆN QUÂN Y

Phạm Minh Đàm<sup>1</sup>, Nguyễn Việt Hưng<sup>1</sup>, Phạm Ngọc Hùng<sup>1</sup>

### TÓM TẮT

**Mục tiêu:** Đánh giá chất lượng các câu hỏi và đề thi trắc nghiệm khách quan của các học phần nội khoa, ngoại khoa theo lý thuyết đáp ứng câu hỏi (IRT: Item Response Theory) bằng phần mềm Conquest. **Đối tượng và phương pháp:** Kết quả thi trắc nghiệm các học phần nội khoa, ngoại khoa được trích xuất từ phần mềm thi trắc nghiệm và xử lý kết quả bằng phần mềm Conquest để đưa ra các thông số của câu hỏi, đề thi trắc nghiệm. **Kết quả:** Đề thi phù hợp với mô hình IRT; 08 đề thi có độ tin cậy cao (Sr - Separation reliability từ 0,86 - 0,96). Các đề thi với nhiều câu hỏi có độ khó nằm ngoài năng lực của thí sinh (quá dễ hoặc quá khó), giá trị độ phân biệt và đáp án nhiễu chưa cao. **Kết luận:** Các đề thi có độ tin cậy cao, tuy nhiên có nhiều câu hỏi cần xem xét điều chỉnh về độ khó, độ phân biệt, đáp án nhiễu trước khi sử dụng làm ngân hàng câu hỏi thi.

\*Từ khóa: Trắc nghiệm; Lý thuyết khảo thí cổ điển; Thuyết đáp ứng câu hỏi; Phần mềm Conquest.

### *Evaluation on the Quality of Questions and Objective Tests of General Surgery and Internal Medicine Modules at the Vietnam Military Medical University*

#### **Summary**

**Objectives:** To evaluate the quality of questions and objective tests of surgery and internal medicine modules by application of Item Response Theory (IRT). **Subjects and methods:** The results of the tests were extracted from the testing data of Vietnam Military Medical University and were analyzed by the Conquest software to provide the parameters of the questions and objective tests. **Results:** Tests fit with IRT model; eight tests achieve high reliability level (Sr - Separation reliability is from 0.86 to 0.96); many test questions are found with difficulty level beyond the ability of candidates (too easy or too difficult); the discriminant value and the noisy answer are not high. **Conclusion:** The test questions have high reliability, but many questions need to be considered and adjusted in terms of difficulty, discriminant, and noisy answers before being used for an exam question bank.

\***Keywords:** Objective tests; Classical test theory; Item Response Theory; Conquest software.

---

<sup>1</sup>Học viện Quân y

Người phản hồi: Nguyễn Việt Hưng (nguyenviethung1884@gmail.com)

Ngày nhận bài: 25/8/2021

Ngày bài báo được đăng: 30/8/2021

## **ĐẶT VẤN ĐỀ**

Trong những năm gần đây, tại Học viện Quân y, bên cạnh việc đổi mới chương trình đào tạo và phương pháp dạy - học thì hoạt động kiểm tra đánh giá cũng có nhiều thay đổi thông qua các hình thức kiểm tra đánh giá mới. Hoạt động kiểm tra đánh giá nhằm cung cấp kịp thời các thông tin về năng lực của người học, chỉ dẫn, tạo động lực cho người học, xây dựng biện pháp quản lý nâng cao chất lượng đào tạo trong tương lai cũng như cung cấp các kết quả về chất lượng đào tạo cho xã hội. Hiện nay, Học viện Quân y đang áp dụng nhiều hình thức kiểm tra đánh giá năng lực người học như trắc nghiệm (MCQs), hỏi - đáp trực tiếp, tự luận, chấm điểm qua bảng kiểm... Các phương pháp kiểm tra đánh giá đều có những ưu, nhược điểm riêng; do vậy, việc áp dụng các phương pháp cần linh hoạt và phù hợp với mục đích của việc kiểm tra đánh giá, nhưng vẫn phải đảm bảo độ tin cậy và độ giá trị để nâng cao tính chính xác trong quá trình đánh giá. Phương pháp đánh giá kiến thức bằng hình thức trắc nghiệm có nhiều ưu điểm và đang được sử dụng rộng rãi ở nhiều bộ môn tại Học viện Quân y. Tuy nhiên, việc đánh giá câu hỏi và đề thi trắc nghiệm chưa được thực hiện. Do đó, để nâng cao chất lượng của câu hỏi trắc nghiệm khách quan cũng như đánh giá mức độ phù hợp của đề thi với năng lực của sinh viên, nhóm nghiên cứu đã tiến hành phân tích dữ liệu kết quả thi để đánh giá chất lượng câu hỏi và đề thi trắc nghiệm khách quan ở môn học Nội khoa và Ngoại khoa nhằm: *Xác định độ tin cậy, độ khó, độ phân biệt của các câu hỏi trắc nghiệm, trên cơ sở đó đề ra khuyến cáo cho các bộ môn.*

## **ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU**

### **1. Đối tượng nghiên cứu.**

Dữ liệu kết quả thi các môn Nội cơ sở, Nội bệnh học, Ngoại cơ sở, Ngoại bệnh học của các lớp DH51A, DH51B, DY16A1, DY16A2, DH52A, DH52B, DH52C, DH52 và DY17 được trích xuất từ phần mềm thi trắc nghiệm. Sau đó, kết quả thi được phân tích bằng phần mềm Conquest để đưa ra các thông số về mức độ phù hợp của câu hỏi với mô hình Rasch, năng lực thí sinh với độ khó của câu hỏi, độ tin cậy của đề thi và các đặc trưng của câu hỏi, đó là độ khó, độ phân biệt, hệ số tương quan của câu hỏi với toàn bài, độ tin cậy và sai số.

### **2. Phương pháp nghiên cứu.**

#### **2.1 Quy trình xây dựng đề thi trắc nghiệm khách quan**

- Bước 1:

Xác định mục tiêu cần đánh giá. Mục tiêu đánh giá cần bám sát mục tiêu của môn học (Chuẩn đầu ra môn học) bao gồm cả nội dung và mức độ cần đánh giá, đồng thời cũng cần chú ý đến mục đích khác của kiểm tra đánh giá như phân loại người học, xét thứ hạng...

- Bước 2:

Xây dựng bảng mô tả đề thi (ma trận đề thi). Căn cứ vào mục tiêu đánh giá, mục tiêu và nội dung bài học, sự đóng góp của bài học và mục tiêu môn học cần xây dựng bảng mô tả đề thi với trọng số phù hợp. Bảng mô tả đề thi được thiết kế theo ma trận hai chiều (nội dung/vấn đề đánh giá, mức độ đánh giá theo thang nhận thức Bloom) để thuận tiện trong quá trình xây dựng nội dung câu hỏi thi trắc nghiệm.

## SỐ ĐẶC BIỆT CHUYÊN ĐỀ VỀ ĐÀO TẠO Y KHOA DỰA TRÊN NĂNG LỰC VÀ CHUẨN ĐẦU RA - 2021

- Bước 3:

Biên soạn câu hỏi trắc nghiệm theo bảng trọng số. Việc biên soạn câu hỏi thi có thể được thực hiện bởi một hoặc nhiều giảng viên theo bản mô tả đề thi. Sau biên soạn, câu hỏi được các chuyên gia phân tích nhằm xác định nội dung và kỹ thuật viết cho từng câu hỏi. Những câu hỏi có nội dung không phù hợp với chương trình dạy học, không rõ ràng... sẽ được loại bỏ hoặc biên tập trước khi sử dụng.

- Bước 4:

Phân tích bộ câu hỏi bằng phương pháp chuyên gia. Sau khi hoàn thành biên soạn xong bộ câu hỏi, các chuyên gia theo lĩnh vực chuyên môn sẽ phân tích, đánh giá, nhận xét góp ý, phản biện từng câu hỏi thi. Những câu hỏi được nhận xét chưa đạt yêu cầu sẽ bị loại bỏ, thay thế hoặc chỉnh sửa lại để đưa vào sử dụng.

- Bước 5:

Thử nghiệm và phân tích kết quả. Các câu hỏi đạt yêu cầu về nội dung và kỹ

thuật viết được sử dụng thi thử và phân tích kết quả. Bước phân tích này chính là phân tích lần hai bằng phương pháp khảo thí hiện đại dựa trên mô hình đáp ứng câu hỏi và phần mềm Conquest. Các câu hỏi đạt chất lượng về thông số kỹ thuật được sử dụng làm ngân hàng câu hỏi thi.

Hiện nay, tại Học viện Quân y, ngày càng có nhiều bộ môn đăng ký kiểm tra đánh giá bằng hình thức trắc nghiệm. Do đó, bước 5 chưa được thực hiện. Ngân hàng câu hỏi thi sau khi được biên soạn, thẩm định về nội dung và kỹ thuật viết sẽ được đưa vào sử dụng để đánh giá học viên. Chính vì thế, việc sử dụng kết quả thi trắc nghiệm để phân tích các thông số của câu hỏi và đề thi trắc nghiệm là cần thiết. Từ kết quả phân tích, chúng tôi đưa ra khuyến nghị với các bộ môn về những câu hỏi có thông số chưa phù hợp, giúp bộ môn có thể điều chỉnh, bổ sung ngân hàng câu hỏi nhằm nâng cao chất lượng và hiệu quả phương pháp kiểm tra đánh giá bằng hình thức trắc nghiệm.

### 2.2 Cấu trúc đề thi

Căn cứ vào chuẩn đầu ra của môn học, nội dung của chương trình dạy học, đề thi của các học phần Nội bệnh học, Nội cơ sở, Ngoại bệnh học, Ngoại cơ sở được thiết kế theo cấu trúc như sau: Đề thi tổng hợp các bộ môn liên quan; số lượng câu hỏi của mỗi bộ môn là 15 câu với tỷ trọng các câu hỏi ở mức nhớ, hiểu, vận dụng lần lượt 50:30:20. Các câu hỏi trong đề thi được chọn tự động bằng phần mềm trong ngân hàng câu hỏi thi.

Bảng 1: Cấu trúc đề thi.

STT	Học phần	Lớp thi	Số đề thi	Số câu hỏi/đề thi	Cấu trúc đề thi		
					Nhớ (%)	Hiểu (%)	Vận dụng (%)
1	Nội bệnh học	16A2, 51B	02	90	50	30	20
2	Ngoại bệnh học	16A1, 51A	02	90	50	30	20
3	Nội cơ sở	52A, 52B	02	105	50	30	20
4	Ngoại cơ sở	17, 52C, 52D	03	105	50	30	20

### **2.3 Lý thuyết khảo thí và phần mềm Conquest**

Hiện nay, để thực hiện đo lường trong giáo dục, các nhà giáo dục vẫn sử dụng hai trường lý thuyết đánh giá cơ bản, đó là: Lý thuyết khảo thí cổ điển (Classical Test Theory - CTT) và lý thuyết đáp ứng câu hỏi hay lý thuyết đánh giá hiện đại (Item Response Theory - IRT). Sự phát triển của lý thuyết khảo thí hiện đại đã khắc phục được những hạn chế của lý thuyết khảo thí cổ điển bằng việc mô hình hóa mối quan hệ giữa năng lực tự tiềm ẩn của thí sinh với xác suất thí sinh trả lời đúng một câu hỏi [4]. Sự đáp ứng của thí sinh với câu hỏi và năng lực của thí sinh được mô tả bằng một hàm đặc trưng của câu hỏi [1] và các tham số đặc trưng của câu hỏi độc lập với mẫu khảo sát [5].

Để đánh giá mối liên quan giữa năng lực của thí sinh với câu hỏi của đề thi trắc nghiệm, nhà Toán học George Rasch đã đưa ra mô hình đáp ứng câu hỏi. Mô hình của Rasch đề cập đến tham số độ khó và năng lực của thí sinh. Theo mô hình đó, thí sinh có năng lực cao hơn thì xác suất trả lời đúng câu hỏi bất kỳ cao hơn thí sinh có năng lực thấp và xác suất trả lời đúng các câu hỏi khó hấp hơn câu hỏi dễ [6].

Phần mềm Conquest là phần mềm được xây dựng trên cơ sở lý thuyết đáp ứng câu hỏi IRT bởi Hội đồng Nghiên cứu Giáo dục Úc, cho phép người dùng tìm hiểu về các thông số trong đánh giá năng lực cũng như đánh giá truyền thống. Kết quả của phần mềm Conquest trong đánh giá đề thi cho chúng ta biết các thông số: Mức độ phù hợp của câu hỏi và đề thi kiểm tra với mô hình Rasch; sự phù hợp của các câu hỏi với nhau; độ tin cậy của đề thi; độ khó, độ phân biệt theo lý thuyết

khảo thí cổ điển; độ khó, độ phân biệt, độ phỏng đoán theo lý thuyết khảo thí hiện đại, giá trị của các đáp án nhiễu.

### **2.4 Các tiêu chí đánh giá câu hỏi và đề thi trắc nghiệm theo phần mềm Conquest.**

#### **2.4.1 Mức độ phù hợp của câu hỏi với mô hình IRT**

Mức độ phù hợp của câu hỏi với mô hình IRT được xác định dựa vào giá trị “sai số bình phương trung bình” (MNSQ - Mean Square) và giá trị “chuẩn hóa Z” (ZSTD - Z standardized). Giá trị MNSQ và ZSTD được phân loại thành 2 chỉ số là infit và outfit. Outfit tính toán độ phù hợp của câu hỏi với mô hình trong đó không sử dụng hệ số đi kèm, infit tính toán sự phù hợp của câu hỏi với mô hình có sử dụng hệ số đi kèm nhằm giảm thiểu ảnh hưởng của những câu hỏi có độ lệch lớn so với những câu hỏi còn lại. Các câu hỏi phù hợp với mô hình là câu hỏi có giá trị MNSQ nằm trong khoảng CI (Confidence Interval), thường trong khoảng (0.7 - 1.3) và giá trị ZSTD nằm trong khoảng (-2; 2) [8].

#### **2.4.2 Độ khó của câu hỏi**

Theo lý thuyết khảo thí cổ điển, độ khó của câu hỏi là tỷ lệ phần trăm thí sinh trả lời đúng câu hỏi đó trong tổng số thí sinh dự thi (tỷ lệ phần trăm chọn đáp án đúng ở từng câu hỏi trong file .ITN). Giá trị của nó có thể chấp nhận được nằm trong khoảng 0,25 - 0,75; câu hỏi có  $p < 0,25$  là quá khó, câu hỏi có  $p > 0,75$  là quá dễ với thí sinh [3]. Theo lý thuyết đáp ứng câu hỏi, độ khó của câu hỏi là xác suất thí sinh trả lời đúng câu hỏi đó. Giá trị độ khó của câu hỏi theo IRT có thể nhận từ  $-\infty$  đến  $+\infty$  và được chia thành 5 mức: rất dễ ( $< -2,0$ ); dễ (từ  $-2,0$  đến  $< -0,5$ ); khó (từ  $-0,5$  đến  $< 2,0$ );

rất khó ( $\geq 2,0$ ) [7]. Tuy nhiên, giá trị độ khó quá thấp hoặc quá cao không có ý nghĩa trong đo lường năng lực của thí sinh dự thi. Theo Baker (2001), giá trị độ khó nên nằm ở mức từ -3 đến +3, các câu hỏi có giá trị nằm ngoài khoảng trên cần xem xét trước khi đưa vào sử dụng [7].

#### **2.4.3 Độ phân biệt của câu hỏi**

Độ phân biệt là khả năng phân biệt giữa nhóm thí sinh có năng lực cao và nhóm thí sinh có năng lực thấp của một câu hỏi trắc nghiệm. Điều đó có nghĩa là tỷ lệ trả lời đúng câu hỏi khó của thí sinh có năng lực cao lớn hơn thí sinh năng lực thấp. Theo lý thuyết khảo thí CTT, giá trị của độ phân biệt rất tốt khi giá trị này  $>0.4$ ; tốt: từ  $0.3 - < 0.4$ ; tạm được: từ  $0.2 - < 0.3$ ; kém:  $< 0.2$  [2]. Những câu hỏi được sử dụng trong đề thi trắc nghiệm nên có giá trị độ phân biệt  $> 0.2$  [4].

#### **2.4.4 Phương án nhiễu**

Phương án gây nhiễu (mồi nhử) là các phương án ngoài đáp án. Phương án nhiễu tốt là phương án có tỷ lệ lựa chọn gần với tỷ lệ mong muốn được tính theo công thức:

$$i = \frac{1-p}{k-1} \times 100\%$$

Trong đó:

i: tỷ lệ mồi nhử mong muốn;

p: độ khó của câu hỏi;

k: tổng số phương án trả lời của câu hỏi.

Cách tính này cho phép xác định mồi nhử không hấp dẫn khi tỷ lệ lựa chọn nhỏ hơn 50% tỷ lệ mong muốn. Câu hỏi thi có chất lượng tốt khi có xác suất lựa chọn các phương án sai (mồi nhử) là tương

đương nhau. Các phương án bị có một số ít (hoặc không có) thí sinh lựa chọn chứng tỏ rằng phương án sai đó là quá lộ liễu, làm tăng khả năng đoán đúng của thí sinh. Những phương án sai có quá nhiều thí sinh lựa chọn có thể những là những phương án gây hiểu nhầm.

#### **2.4.5 Độ tin cậy của đề thi**

Độ tin cậy của đề thi trắc nghiệm chính là đại lượng biểu thị mức độ chính xác của phép đo nhờ đề thi trắc nghiệm. Có nhiều cách xác định độ tin cậy của đề thi như: trắc nghiệm - trắc nghiệm lại; phân đôi đề thi trắc nghiệm; các đề thi trắc nghiệm tương đương; Kuder-Richardson và sử dụng hệ số Cronbach Alpha. Lý thuyết IRT sử dụng hệ số Cronbach's Alpha trong đánh giá độ tin cậy của đề thi. Độ tin cậy của đề thi là tốt khi giá trị Cronbach's Alpha lớn hơn hoặc bằng 0.8 [8].

#### **2.4.6 Độ giá trị của đề thi**

Yêu cầu quan trọng nhất của đề thi trắc nghiệm với tư cách là công cụ đo lường trong giáo dục đó là đo đúng nội dung cần đo. Để đạt được độ giá trị cao, cần bám sát mục tiêu cần đo lường trong quá trình xây dựng ngân hàng câu hỏi thi cũng như quá trình tổ chức và triển khai kỳ thi. Nếu thực hiện các quá trình nói trên không đúng thì khả năng kết quả của sự đo lường sẽ không phản ánh đúng mục tiêu cần đo [1]. Qua xem xét độ tin cậy và độ giá trị của đề thi cho thấy, đề thi trắc nghiệm không có độ tin cậy thì chúng ta không thể nói đến độ giá trị. Tuy nhiên, một đề thi có độ tin cậy cao, chưa chắc đã đạt được độ giá trị bởi đề thi không đo đúng mục tiêu cần đo lường. Do đó, một đề thi trắc nghiệm cần đạt được độ tin cậy cao, và đo đúng nội dung cần đo (độ giá trị).

**KẾT QUẢ NGHIÊN CỨU VÀ BÀN LUẬN**

**1. Mức độ phù hợp mô hình IRT**

Mức độ phù hợp của câu hỏi với mô hình được thể hiện trong bảng kết quả SHW. Kết quả phân tích cho thấy, ở các đề thi mặc dù có nhiều câu hỏi có giá trị UNWEIGHTED FIT nằm ngoài khoảng tin cậy tương ứng nhưng giá trị WEIGHTED FID vẫn nằm trong khoảng tin cậy CI, giá trị T nằm trong khoảng [-2;2]. Do đó, các câu hỏi vẫn được đánh giá phù hợp với mô hình IRT đang được xem xét. Các đề thi có có hệ số tin cậy cao từ 0,86 - 0,96.

*Bảng 2: Kết quả kiểm định sự phù hợp của câu hỏi với mô hình.*

	Nội khoa				Ngoại khoa				
	Bệnh học		Cơ sở		Bệnh học		Cơ sở		
	16A2	51B	52A	52B	16A1	51A	52C	52D	17
Điểm thi trung bình	6,91	6,96	6,44	5,87	8,84	9,42	7,17	7,85	8,32
Sr	0,96	0,96	0,92	0,93	0,86	0,67	0,92	0,91	0,87
Chi-square	2620	2680	1139	1060	3935	4759	1629	2242	2146
df	90	88	103	102	89	86	88	90	87
Sig	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

**2. Đặc tính của các câu hỏi trong đề thi**

Câu hỏi có chất lượng tốt được sử dụng vào ngân hàng câu hỏi và phục vụ cho các kỳ thi là những câu hỏi có nội dung phù hợp với nội dung giảng dạy, giá trị độ khó, độ phân biệt và các phương án nhiễu hợp lý. Các câu hỏi không phù hợp với mô hình, quá dễ, quá khó, không có độ phân biệt và các phương án nhiễu không hiệu quả thì cần có sự điều chỉnh để sử dụng hoặc loại bỏ khỏi ngân hàng câu hỏi.

**\* Độ khó:**

Phần mềm Conquest cho người đọc thông tin độ khó của câu hỏi theo cả CTT và IRT. Kết quả độ khó của câu hỏi được thể hiện ở 2 file .SHW và .ITN (giá trị ItermTresold). Kết quả phân tích cho thấy, các đề thi học phần của bộ môn Nội khoa có nhiều câu hỏi ở mức độ khó và quá khó hơn so với các đề thi học phần Ngoại khoa. Bên cạnh đó, tỷ lệ câu hỏi ở các nhóm giữa các đề thi trong cùng học phần của môn học cũng không giống nhau.

## SỐ ĐẶC BIỆT CHUYÊN ĐỀ VỀ ĐÀO TẠO Y KHOA DỰA TRÊN NĂNG LỰC VÀ CHUẨN ĐẦU RA - 2021

*Bảng 3:* Kết quả tổng hợp độ khó của câu hỏi.

	Nội khoa				Ngoại khoa				
	Bệnh học		Cơ sở		Bệnh học		Cơ sở		
	16A2	51B	52A	52B	16A1	51A	52C	52D	17
Quá khó	04	01	03	04	0	0	0	0	0
Khó	27	24	48	54	05	0	20	11	13
Dễ	24	43	49	40	18	1	49	42	28
Quá dễ	35	21	03	05	67	86	21	37	47
Tổng số CH	90	89	104	103	90	87	89	90	88

\* *Độ phân biệt:*

Độ phân biệt của câu hỏi được thể hiện bằng chỉ số Discrimination. Kết quả tổng hợp câu hỏi đạt yêu cầu theo giá trị độ phân biệt được thể hiện trong Bảng 4.

*Bảng 4:* Kết quả tổng hợp độ phân biệt của câu hỏi.

Độ phân biệt	Nội khoa				Ngoại khoa				
	Bệnh học		Cơ sở		Bệnh học		Cơ sở		
	16A2	51B	52A	52B	16A1	51A	52C	52D	17
< 0,2	20	27	39	50	17	14	24	29	26
≥ 0,2	70	62	65	53	73	73	65	61	62
Tổng số CH	90	89	104	103	90	87	89	90	88

Kết quả cho thấy, trong các đề thi, số câu hỏi có độ phân biệt chấp nhận được chiếm tỷ lệ cao nhưng cần xem xét đến các thông số khác của câu hỏi khi sử dụng làm ngân hàng đề thi.

\* *Phương án nhiễu:*

Kết quả tổng hợp chất lượng phương án nhiễu cho thấy, tỷ lệ các câu hỏi có chất lượng phương án nhiễu không hiệu quả ở các đề thi dao động từ 42,4% đến 78,16%. Do đó, các đề thi cần xem xét, điều chỉnh nâng cao chất lượng phương án nhiễu góp phần nâng cao chất lượng câu hỏi thi. Kết quả cụ thể được thể hiện qua Bảng 5.

*Bảng 5:* Kết quả câu hỏi có chất lượng phương án nhiễu không hiệu quả.

	Nội khoa				Ngoại khoa				
	Bệnh học		Cơ sở		Bệnh học		Cơ sở		
	16A2	51B	52A	52B	16A1	51A	52C	52D	17
Tổng số CH	90	89	104	103	90	87	89	90	88
Câu hỏi có phương án nhiễu không hiệu quả	38 (42,4%)	37 (41,57%)	36 (43,62%)	44 (42,72%)	65 (72,22%)	68 (78,16%)	45 (50,56%)	62 (68,89%)	63 (71,59)

## SỐ ĐẶC BIỆT CHUYÊN ĐỀ VỀ ĐÀO TẠO Y KHOA DỰA TRÊN NĂNG LỰC VÀ CHUẨN ĐẦU RA - 2021

### 3. Phân tích thông số của một số câu hỏi trắc nghiệm

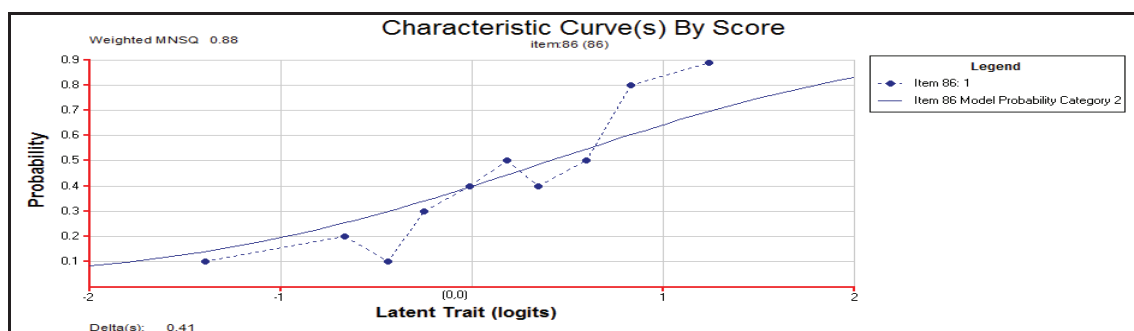
Dựa vào kết quả phân tích độ giá trị (nội dung của câu hỏi), độ khó, độ phân biệt, giá trị của các phương án nhiễu để xác định các nhóm câu hỏi có chất lượng tốt, nhóm câu hỏi cần chỉnh sửa và nhóm câu hỏi cần xem xét trước khi sử dụng hoặc loại bỏ.

*\* Nhóm câu hỏi có chất lượng tốt:*

Có thể sử dụng trong ngân hàng câu hỏi thi là những câu hỏi có nội dung rõ ràng phù hợp với nội dung cần kiểm tra đánh giá, có độ khó phù hợp, độ phân

biệt tốt và các đáp án nhiễu có giá trị. Kết quả phân tích câu hỏi số 86 đề thi Nội bệnh học lớp 16A2 cho thấy: mức độ phù hợp của câu hỏi là 0,88 (Weighted MNSQ); về nội dung câu hỏi dùng để đánh giá nhận thức của sinh viên ở mức hiểu; có độ khó 0,41 (Độ khó theo IRT: Item Threshold là 0.41); độ phân biệt (Discrimination) ở mức 0,45 là phù hợp để phân loại sinh viên; các đáp án nhiễu đều có giá trị hợp lý. Kết quả phân tích đường cong đặc trưng của câu hỏi cũng cho thấy, năng lực của sinh viên bám sát đường cong kỳ vọng.

Label	Score	Count	% of tot	Pt Bis	t	(p)	WLEAvg:1	WLE
A	0.00	13	13.13	-0.19	-1.86	(.066)	-0.38	0.79
B	0.00	26	26.26	-0.39	-4.16	(.000)	-0.53	0.88
C	0.00	19	19.19	0.03	0.26	(.799)	0.01	0.46
D	1.00	41	41.41	0.45	5.02	(.000)	0.46	0.64



Hình 1: Các thông số và hàm thông tin của câu hỏi số 86 đề thi Nội bệnh học lớp 16A2.

*\* Nhóm câu hỏi cần xem xét trước khi sử dụng hoặc loại bỏ:*

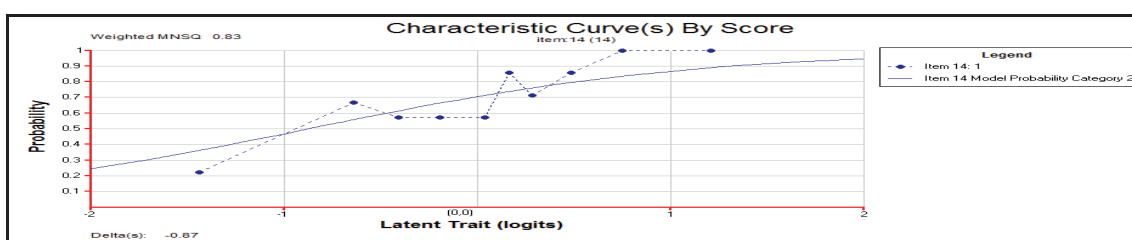
Nhóm câu hỏi có nội dung chưa rõ ràng, gây nhầm lẫn cho thí sinh hoặc không phù hợp với mục tiêu kiểm tra đánh giá; độ khó, độ phân biệt chưa đảm bảo hoặc các đáp án nhiễu chưa hiệu quả là các nhóm câu hỏi có thể sử dụng nhưng cần được



## SỐ ĐẶC BIỆT CHUYÊN ĐỀ VỀ ĐÀO TẠO Y KHOA DỰA TRÊN NĂNG LỰC VÀ CHUẨN ĐẦU RA - 2021

điều chỉnh để các thông số câu hỏi hợp lý. Phân tích câu hỏi số 14 của đề thi Ngoại cơ sở lớp 52C cho thấy:

Label	Score	Count	% of tot	Pt Bis	t	(p)	WLEAvg:1	WLE
A	0.00	18	25.35	-0.55	-5.46	(.000)	-0.70	0.80
B	1.00	49	69.01	0.57	5.79	(.000)	0.29	0.57
C	0.00	4	5.63	-0.11	-0.92	(.359)	-0.39	0.34



Hình 2: Các thông số và hàm thông tin của câu hỏi số 14 đề thi Ngoại cơ sở lớp 52C.

Về nội dung, câu hỏi dùng để đánh giá kiến thức ở mức độ nhớ, với độ khó 0,69, độ phân biệt là 0,57 cùng với đồ thị thông tin về câu hỏi là hợp lý. Tuy nhiên, khi phân tích thông số các đáp án có thể thấy, trong các đáp án không có sinh viên nào chọn đáp án D. Điều đó chứng tỏ đáp án D không có giá trị làm nhiều. Câu hỏi trên nếu được điều chỉnh các phương án trả lời thì có thể sử dụng trong ngân hàng câu hỏi thi.

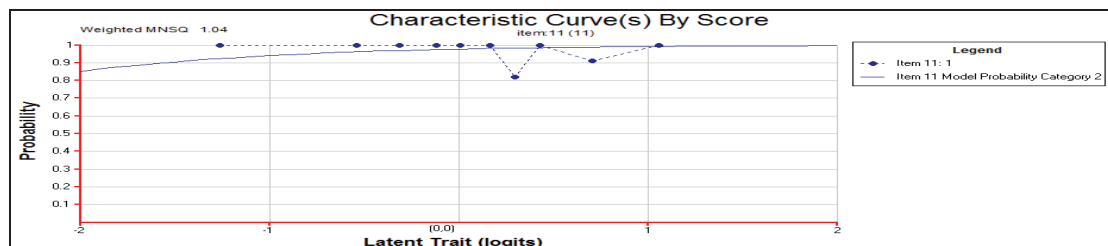
Nhóm các câu hỏi không đáp ứng được cần loại bỏ hoặc điều chỉnh nhiều, đó là các câu hỏi không có độ phân biệt (độ phân biệt bằng 0 hay thậm chí âm); độ khó bằng 1 hoặc bằng 0. Phân tích câu hỏi số 38 đề thi Ngoại bệnh học lớp 16A1 cho thấy nội dung câu hỏi dùng để đánh giá kiến thức ở mức độ nhớ. Tuy nhiên, độ khó của câu hỏi là 1 (100% số sinh viên trả lời đúng), độ phân biệt là 0 (không phân biệt được các nhóm sinh viên).

Label	Score	Count	% of tot	Pt Bis	t	(p)	WLEAvg:1	WLE
A	1.00	104	100.00	NA	NA	(.000)	-0.09	1.07

Hình 3: Thông số câu hỏi 38 đề thi Ngoại bệnh học 16A1.

Khi phân tích hàm thông tin câu hỏi số 11 đề thi Nội bệnh học của lớp 51B cho thấy xác suất thí sinh có năng lực thấp trả lời đúng câu hỏi cao hơn các thí sinh có năng lực cao (tức độ phân biệt âm).

## SỐ ĐẶC BIỆT CHUYÊN ĐỀ VỀ ĐÀO TẠO Y KHOA DỰA TRÊN NĂNG LỰC VÀ CHUẨN ĐẦU RA - 2021



Hình 4: Hàm thông tin câu hỏi 11 đề thi Nội bệnh học lớp 51B.

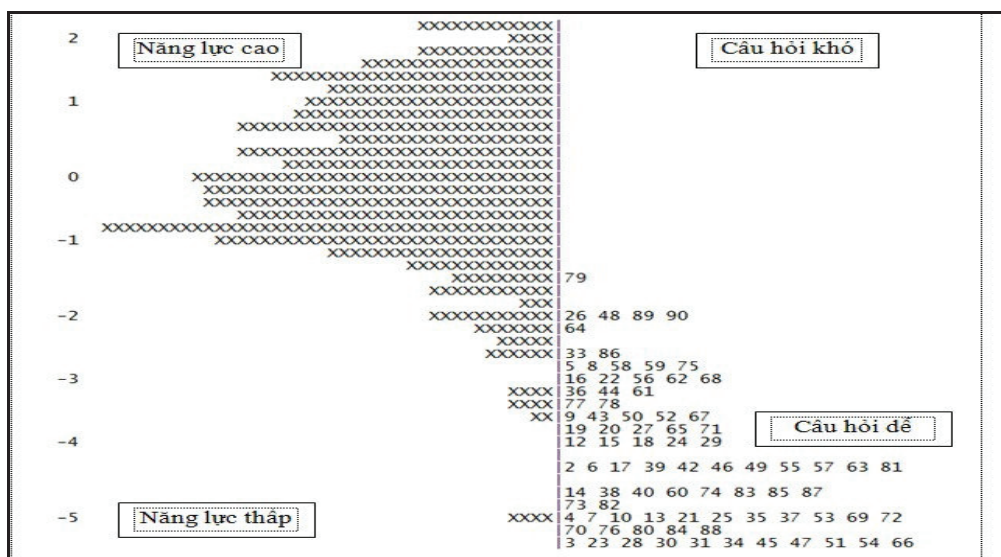
Bảng 6: Kết quả chất lượng câu hỏi.

Nhóm câu hỏi	Nội khoa				Ngoại khoa				
	Bệnh học		Cơ sở		Bệnh học		Cơ sở		
	16A2	51B	52A	52B	16A1	51A	52C	52D	17
Chất lượng tốt	21	17	31	18	03	0	7	2	6
Cần xem xét chỉnh sửa	69	72	74	85	87	87	82	98	82
Tổng	90	89	104	103	90	87	89	90	88

Các câu hỏi sau khi được xem xét đến các thông số như độ phù hợp mô hình, độ khó, độ phân biệt, giá trị các đáp án nhiều cho thấy tỷ lệ các câu hỏi cần xem xét chỉnh sửa trước khi sử dụng chiếm số lượng lớn từ 71,15% - 100%. Bên cạnh đó, số câu hỏi có chất lượng tốt ở các đề thi thuộc học phần môn Nội khoa nhiều hơn ở các đề thi môn Ngoại khoa.

### 4. Phân bố độ khó của câu hỏi với năng lực của thí sinh ở các đề thi

Kết quả phân tích độ khó của câu hỏi với năng lực của thí sinh được thể hiện trong bảng kết quả .SHW cho thấy hầu hết các đề thi có nhiều câu hỏi có độ khó nằm ngoài năng lực của sinh viên.



Hình 5: Phân bố độ khó của câu hỏi với năng lực của sinh viên của đề thi lớp 51A

## SỐ ĐẶC BIỆT CHUYÊN ĐỀ VỀ ĐÀO TẠO Y KHOA DỰA TRÊN NĂNG LỰC VÀ CHUẨN ĐẦU RA - 2021

Đề thi đánh giá năng lực của sinh viên từ < -5 đến -1, số lượng câu hỏi thi tập trung đánh giá thí sinh có năng lực trung bình và thấp. Đề thi có nhiều thí sinh có năng lực cao mà không có đánh giá. Ngược lại, đề thi có nhiều câu hỏi mà năng lực thí sinh vượt quá độ khó của các câu hỏi này.

Bảng 6: Kết quả phân tích đề thi.

Câu hỏi đánh giá	Nội khoa				Ngoại khoa				
	Bệnh học		Cơ sở		Bệnh học		Cơ sở		
	16A2	51B	52A	52B	16A1	51A	52C	52D	17
Khoảng năng lực đánh giá	-3,0 đến 3,0	-2,5 đến 2,5	-2,0 đến 2,0	-1.5 đến 1,0	-4,5 đến 0,5	-5,5 đến - 2,0	-2,5 đến 2,5	-2,5 đến 1,5	-3,0 đến 2,0
Khoảng năng lực của SV	-2 đến 2	-2 đến 1	-1,5 đến 1,5	-0.5 đến 1	-2,5 đến 2,5	-3 đến 3	-2 đến 1,5	-2 đến 2	-2 đến 2
Đánh giá nhóm năng lực cao	15 (06 câu quá khó)	11 (06 câu quá khó)	16 (06 câu quá khó)	29 (17 câu quá khó)	0	0	05	03	02
Đánh giá nhóm năng lực trung bình	31	34	45	40	22	1	65	23	31
Đánh giá nhóm năng lực thấp và rất thấp	44	44	33	34	68	86	20	64	55
Tổng	90	89	104	103	90	87	89	90	88

Kết quả phân tích biểu đồ phân bố năng lực của thí sinh với độ khó của câu hỏi cho thấy các đề thi học phần Nội khoa có nhiều câu hỏi đánh giá thí sinh có năng lực cao hơn ở các đề thi Ngoại khoa. Tuy nhiên, trong các đề thi Nội khoa có những câu hỏi quá khó, vượt quá năng lực trả lời của thí sinh. Bên cạnh đó, biểu đồ phân bố năng lực của thí sinh với độ khó của câu hỏi cũng cho thấy các đề thi có nhiều câu hỏi ở mức quá dễ mà năng lực của thí sinh vượt quá độ khó của các câu hỏi này.

Qua biểu đồ phân bố cho thấy các đề thi học phần Nội khoa phù hợp để đánh giá thí sinh, tuy nhiên cần xem xét lại những câu hỏi quá dễ hoặc quá khó nằm ngoài khả năng của thí sinh. Các đề thi học phần Ngoại khoa có nhiều câu hỏi có độ khó thấp hơn so với năng lực của thí sinh.

### **KẾT LUẬN**

Phân tích kết quả thi trắc nghiệm khách quan của các đề thi học phần Nội khoa và Ngoại khoa bằng phần mềm Conquest cho thấy: Các đề thi tương đối phù hợp với mô hình IRT, có độ tin cậy và giá trị cao.

Tuy nhiên, trong từng đề thi còn có nhiều câu hỏi mà độ giá trị, độ phân biệt chưa cao. Mặt khác, giá trị của các đáp án nhiều ở nhiều câu hỏi chưa tốt, chưa phát huy được khả năng gây nhiễu. Qua việc phân tích đề thi kiểm tra của các học phần Ngoại khoa và Nội khoa bằng phần mềm Conquest, có thể lựa chọn được những câu hỏi tốt, đề xuất những câu hỏi cần chỉnh sửa và loại bỏ những câu hỏi không chất lượng, từ đó nâng cao chất lượng ngân hàng câu hỏi, góp phần đánh giá đúng năng lực nhận thức của người học.

### **TÀI LIỆU THAM KHẢO**

1. Lâm Quang Thiệp. Đo lường và đánh giá hoạt động học tập trong nhà trường. Nhà Xuất bản Đại học Sư phạm, Hà Nội 2012.
2. Duong, T.T. Trắc nghiệm và đo lường thành quả học tập [Test and measure academic achievement]. Hanoi: Social Sciences Publishing House 2005.
3. Lord F.M. Tài liệu dịch: Psychometrika - Mối quan hệ giữa độ tin cậy của câu hỏi đa lựa chọn và phân bố độ khó của câu hỏi 1952; 18:181-194.
4. Harris D. Comparison of 1-, 2-, and 3-paramater ITR models. A module in NCME series of Instructional Topics in Educational Measurement. NCME Journal of Educational Measurement: Issues and Practices 1989; 35-41.
5. Hambleton, R.K., Swaminathan, H. Item response theory: Principles and applications. Springer Science & Business Media 2013.
6. Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danish Institute for Educational Research 1960.
7. Baker, F.B. The basic of item response theory. College Park, MD: University of Maryland, ERIC Clearinghouse on Assessment and Evaluation 2001.
8. Griffin, J. P. (1997), An introduction to the RASCH model, Australia: University of Melbourne