



ĐỊNH CỠ ĐỀ TRẮC NGHIỆM DỰA TRÊN LÝ THUYẾT TRẮC NGHIỆM CỔ ĐIỂN VÀ LÝ THUYẾT ỨNG ĐÁP CÂU HỎI

Phạm Minh Hoàn

Viện Công nghệ Thông tin Kinh tế (SITE), Trường Đại học Kinh tế Quốc dân

Ngày tòa soạn nhận được bài báo: 26/02/2018

Ngày phân biên đánh giá và sửa chữa: 05/03/2018

Ngày bài báo được chấp nhận đăng: 15/03/2018

Tóm tắt:

Bài báo trình bày các khái niệm cơ bản của lý thuyết trắc nghiệm cổ điển và lý thuyết ứng đáp câu hỏi. Trên cơ sở lý thuyết đó, tác giả trình bày quy trình định cỡ đề trắc nghiệm dựa trên hai lý thuyết này, làm cơ sở để hiệu chỉnh giá trị của ngân hàng câu hỏi trắc nghiệm. Kết quả nghiên cứu được thử nghiệm trên phần mềm dưới dạng một WebSite bằng ngôn ngữ C#. Phần mềm cho phép phân tích, đánh giá câu hỏi trắc nghiệm để hỗ trợ định cỡ đề trắc nghiệm và hiệu chỉnh ngân hàng câu hỏi. Nghiên cứu này với mong muốn tiếp tục hoàn thiện phần mềm như một công cụ hỗ trợ đầy đủ cho việc lập ngân hàng câu hỏi, định cỡ đề trắc nghiệm cho các kỳ thi trắc nghiệm.

Từ khóa: Trắc nghiệm cổ điển, lý thuyết ứng đáp câu hỏi, ngân hàng đề thi, đề trắc nghiệm, hàm đặc trưng câu hỏi, Rasch.

1. Đặt vấn đề

Đánh giá kết quả là một công việc bắt buộc để xem xét tính hiệu quả của một công việc nói chung, cho phép lựa chọn phương pháp tiên tiến hơn để thực hiện công việc cũng như năng lực thực hiện công việc đó. Đánh giá kết quả công việc được thực hiện bởi phép đo lường dựa trên quy chuẩn nào đó.

Trong giáo dục và đào tạo, phép đo lường cho phép xác định mục tiêu giáo dục được đặt ra có phù hợp, có đạt được mục tiêu đề ra, việc giảng dạy có thành công, người học có tiếp thu được hay không [1]. Một trong những phương pháp để thực hiện phép đo này là thi trắc nghiệm. Phổ biến hiện nay ở Việt Nam có hai lý thuyết đã được nghiên cứu và áp dụng để phân tích và đánh giá câu hỏi trắc nghiệm là lý thuyết trắc nghiệm cổ điển và lý thuyết ứng đáp câu hỏi. Tuy nhiên, cho đến nay vẫn chưa có công cụ phổ biến nào cho phép hiệu chỉnh giá trị của ngân hàng câu hỏi, định cỡ đề trắc nghiệm.

Với mong muốn có một nghiên cứu về quy trình định cỡ đề trắc nghiệm và hiệu chỉnh ngân hàng đề thi, tác giả đã chọn hướng nghiên cứu “Định cỡ đề trắc nghiệm dựa trên lý thuyết trắc nghiệm cổ điển và lý thuyết ứng đáp câu hỏi”. Phần sau đây trình bày kết quả thử nghiệm trên một đề trắc nghiệm cụ thể cấp chứng chỉ “Ứng dụng Công nghệ Thông tin cơ bản”.

2. Phân tích, đánh giá đề trắc nghiệm dựa trên lý thuyết trắc nghiệm cổ điển

Lý thuyết trắc nghiệm cổ điển (Classical Test Theory - CTT) [2] là một trong những lý thuyết

quan trọng liên quan đến khoa học đo lường được xây dựng dựa trên lý thuyết xác suất và thống kê. Nguyên lý cơ bản của lý thuyết này là phân tích câu hỏi trắc nghiệm nhằm nâng cao chất lượng của chúng, loại bỏ, sửa chữa và tuyển chọn câu hỏi theo yêu cầu.

2.1. Các tham số đặc trưng

• Độ khó

Xác định độ khó dựa vào việc thử nghiệm câu hỏi trắc nghiệm trên các đối tượng thí sinh phù hợp. Độ khó p bằng tỷ số phần trăm thí sinh làm đúng câu hỏi trên tổng số thí sinh tham gia trả lời câu hỏi đó:

$$p = \frac{\text{tổng số thí sinh trả lời đúng}}{\text{tổng số thí sinh trả lời}} \quad (2.1)$$

Việc sử dụng trị số thống kê p để đo độ khó như trên cho biết mức độ khó dễ của các câu hỏi mà không cần xem xét nội dung của chúng. Độ khó của câu hỏi phụ thuộc lĩnh vực khoa học, các lĩnh vực khác nhau có độ khó đặc trưng khác nhau.

• Độ phân biệt

Độ phân biệt là khả năng đề trắc nghiệm phân biệt được năng lực của thí sinh khi tham gia trả lời câu hỏi trắc nghiệm. Lý thuyết trắc nghiệm cổ điển có hai phương pháp tính độ phân biệt:

- Phương pháp tính theo công thức hệ số tương quan Pearson (Pearson product-moment correlation) đo độ tương quan giữa điểm của câu hỏi X và tổng điểm bài trắc nghiệm Y .

$$\rho_{x,y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad (2.2)$$

Giá trị $\rho_{x,y}$ ($-1 \leq \rho_{x,y} \leq 1$) càng lớn nếu câu hỏi có độ phân biệt càng lớn.

- Phương pháp cô điển để tính độ phân biệt dựa vào tổng điểm thô của từng thí sinh, tách thí sinh ra thành hai nhóm: một nhóm giỏi gồm 27% thí sinh đạt điểm cao từ trên xuống, và nhóm kém gồm 27% thí sinh đạt điểm kém từ dưới lên. Gọi C là số thí sinh làm đúng câu hỏi thuộc nhóm giỏi, T là số thí sinh làm đúng câu hỏi thuộc nhóm kém, S là số lượng thí sinh của một trong hai nhóm nói trên, biểu thức tính độ phân biệt D của câu hỏi [8] như sau:

$$D = \frac{C-T}{S} \quad (2.3)$$

• Độ tin cậy

Trắc nghiệm là một phép đo: dùng thước đo là đề trắc nghiệm để đo lường một năng lực nào đó của thí sinh. Độ tin cậy của đề trắc nghiệm chính là đại lượng biểu thị mức độ chính xác của phép đo nhờ đề trắc nghiệm. Để tính độ tin cậy, có hai phương pháp được sử dụng:

- Phương pháp phân đôi đề trắc nghiệm:

Phương pháp phân đôi đề trắc nghiệm thực chất là tạo hai đề trắc nghiệm tương đương, mỗi đề là một nửa của đề trắc nghiệm chung, một nửa đề trắc nghiệm gồm các câu có số thứ tự chẵn và nửa còn lại gồm các câu lẻ. Độ tin cậy nửa đề trắc nghiệm bằng hệ số tương quan giữa hai bộ điểm của hai nửa đề trắc nghiệm, và được tính theo công thức Spearman – Brown:

$$r_2 = \frac{2r_s}{r_s + 1} \quad (2.4)$$

Trong đó: r_s – độ tin cậy của đề trắc nghiệm ngắn xuất phát,

r_2 – độ tin cậy của đề trắc nghiệm có độ dài gấp 2 lần.

- Phương pháp Kuder – Richardson:

Phương pháp Kuder – Richardson tính độ tin cậy dựa trên ý tưởng xem mỗi câu trong đề trắc nghiệm là một đề trắc nghiệm tương đương, tức là chúng có cùng điểm trung bình và cùng phương sai. Công thức Kuder – Richardson ($K-R$) có dạng như sau [8]:

$$r = \frac{k}{k-1} \left(1 - \frac{\sum pq}{\delta^2} \right) \quad (2.5)$$

Trong đó: k – số câu hỏi của đề trắc nghiệm.

p – tỷ lệ trường hợp trả lời đúng cho một câu hỏi.

$q = (1-p)$ – tỷ lệ trường hợp trả lời sai cho một câu hỏi.

δ^2 – phương sai của tổng điểm mọi thí sinh đối với cả đề trắc nghiệm.

Trong trường hợp độ khó của các câu hỏi không khác nhau nhiều, công thức $K-R$ có thể biến

đổi thành công thức Kuder – Richardson rút gọn dễ tính toán hơn [8]:

$$r = \frac{k}{k-1} \left[1 - \frac{M(1 - \frac{M}{k})}{\delta^2} \right] \quad (2.6)$$

Trong đó M là giá trị trung bình điểm số của cả đề trắc nghiệm.

2.2. Phân tích, đánh giá một đề trắc nghiệm

Đánh giá đề trắc nghiệm được bắt đầu bằng phân tích các câu hỏi trong đề đó.

• Phân tích các câu hỏi trắc nghiệm

Để hoàn thiện đề trắc nghiệm cần phải triển khai các trắc nghiệm thử. Trắc nghiệm thử là một phép đo kép: dùng đề trắc nghiệm để đo năng lực thí sinh, đồng thời sử dụng một nhóm thí sinh như một thước đo để đo chất lượng các câu hỏi và đề trắc nghiệm.

Tính tham số độ khó bởi công thức (2.1), các câu hỏi của một đề trắc nghiệm thường phải có độ khó khác nhau. Theo công thức trên thì giá trị p càng bé thì câu hỏi càng khó và ngược lại. Thông thường độ khó của một câu hỏi có thể chấp nhận được nằm trong khoảng 0.25 – 0.75; câu hỏi có độ khó lớn hơn 0.75 là quá dễ, có độ khó nhỏ hơn 0.25 là quá khó.

Tham số độ phân biệt được tính bởi công thức (2.3), những câu hỏi có độ phân biệt nhỏ hơn không hoặc gần bằng không là những câu hỏi có độ phân biệt kém cần phải được chỉnh sửa hoặc loại bỏ khỏi đề trắc nghiệm.

• Tính độ tin cậy của đề trắc nghiệm

Tham số độ tin cậy của đề trắc nghiệm được tính bởi phương pháp phân đôi đề trắc nghiệm bởi công thức (2.4) hoặc công thức Kuder – Richardson bởi công thức (2.5) và (2.6).

3. Định cỡ đề trắc nghiệm dựa trên Lý thuyết ứng đáp câu hỏi trắc nghiệm

Mô hình “ứng đáp câu hỏi” (Item Response Theory - IRT) [2] để mô tả mối tương tác nguyên tố giữa một thí sinh với một câu hỏi của đề trắc nghiệm và dùng mô hình đó để phân tích các dữ liệu thu được từ đề trắc nghiệm.

3.1. Các mô hình

• Mô hình một tham số - Rasch

Mô hình một tham số Rasch biểu diễn quan hệ giữa năng lực thí sinh và độ khó câu hỏi và được tính bởi hàm đặc trưng câu hỏi một tham số [3][4][6][7]:

$$P(\theta) = \frac{e^{\theta-b}}{1 + e^{\theta-b}} \quad (3.1)$$

Trong đó: P – xác suất trả lời đúng câu hỏi,

b – độ khó,

θ – năng lực của thí sinh.

• **Mô hình hai tham số**

Mô hình hai tham số biểu diễn quan hệ giữa năng lực thí sinh và độ khó, độ phân biệt câu hỏi và được tính bởi hàm đặc trưng câu hỏi một tham số [2][3][7]:

$$P(\theta) = \frac{e^{a(\theta-b)}}{1 + e^{a(\theta-b)}} \quad (3.2)$$

Trong đó: a – độ phân biệt.

• **Mô hình ba tham số**

Mô hình ba tham số đưa thêm tham số c (tham số đoán mò), khi đó quan hệ giữa năng lực thí sinh và độ khó, độ phân biệt câu hỏi và tham số đoán mò của thí sinh được tính bởi hàm đặc trưng câu hỏi [2][3][7]:

$$P(\theta) = c + (1 - c) \frac{e^{a(\theta-b)}}{1 + e^{a(\theta-b)}} \quad (3.3)$$

Trong đó: c - tham số đoán mò, khi $\theta \rightarrow 0$ thì $P(\theta) \rightarrow c$.

3.2. Định cỡ đề trắc nghiệm

Để định cỡ đề trắc nghiệm sử dụng mô hình IRT, trước hết giả định các câu hỏi là độc lập với nhau, cần tính xác suất trả lời đúng câu hỏi. Xác suất này phụ thuộc vào năng lực θ của thí sinh và các tham số đặc trưng của câu hỏi. Tuy nhiên, cả hai loại tham số, năng lực thí sinh và đặc trưng của câu hỏi đều chưa biết, cái duy nhất biết được là việc trả lời (ứng đáp) các câu hỏi của thí sinh. Việc xác định các giá trị tham số năng lực θ của từng thí sinh và các tham số a, b, c của từng câu hỏi là quá trình ước lượng từ các kết quả ứng đáp câu hỏi.

Do đó, kỹ thuật định cỡ đề trắc nghiệm gồm 2 giai đoạn: giai đoạn đầu ước lượng các tham số của từng câu hỏi, giai đoạn hai ước lượng các tham số năng lực của từng thí sinh. Hai giai đoạn được thực hiện tương tác với nhau cho đến khi thu được một tập hợp ổn định các tham số ước lượng được.

• **Quá trình ước lượng tham số:**

Giai đoạn thứ nhất, giả sử các ước lượng năng lực của mỗi thí sinh được xem như là đã được biểu diễn trên một thang đo với đơn vị đo thực. Khi đó, các tham số của mỗi câu hỏi được ước lượng theo quy trình ước lượng biến cố hợp lý cực đại [5] [8] (hoặc ước lượng *OLS* - Ordinary Least Square – Phương pháp bình phương nhỏ nhất [4]). Quy trình được thực hiện theo từng câu hỏi một, cuối cùng thu được một bộ các tham số của các câu hỏi trong đề trắc nghiệm.

Giai đoạn thứ hai, giả thiết rằng các giá trị tham số câu hỏi ước lượng được ở giai đoạn trước là các giá trị thực của tham số. Khi đó, năng lực của mỗi thí sinh được ước lượng theo quy trình ước lượng biến cố hợp lý cực đại [5]. Giai đoạn này

được thực hiện để ước lượng năng lực theo từng thí sinh một.

Hai giai đoạn được lặp đi lặp lại cho đến khi thỏa mãn một tiêu chí hội tụ nào đó (các giá trị tham số gần như không thay đổi). Như vậy, các tham số của câu hỏi của đề trắc nghiệm và các năng lực của thí sinh được ước lượng đồng thời. Cách tiệm cận thông minh đó làm cho một bài toán ước lượng rất phức tạp có thể hạ xuống mức có thể giải trên một máy tính. Nội dung của các kỹ thuật sử dụng trong quá trình này sẽ được trình bày trong một bài báo khác.

4. Thử nghiệm và giá kết quả

4.1. Xây dựng chương trình thử nghiệm

Chương trình CTC (CLOUD TEST CREATOR) được xây dựng dưới dạng một WebSite bằng ngôn ngữ lập trình C# trên nền tảng .NET Framework 4.6.1 với chức năng chính là phân tích, đánh giá bộ đề trắc nghiệm. Kết quả phân tích và đánh giá câu hỏi trắc nghiệm sẽ hỗ trợ cho việc lựa chọn các câu hỏi trắc nghiệm cho các kỳ thi sau. Dữ liệu thử nghiệm là bộ đề thi Ứng dụng Công nghệ Thông tin cơ bản, và sẽ áp dụng cho các môn khác nhau khi hoàn thiện.

• **Phân tích câu hỏi**

Câu hỏi của đề trắc nghiệm được phân tích bởi một trong hai lý thuyết:

- Trắc nghiệm cổ điển.

- Ứng đáp câu hỏi: Mô hình 1 tham số (*Rasch*) và mô hình 2 tham số.

• **Chức năng nhập dữ liệu từ bộ kết quả thi trắc nghiệm**

Để nhập kết quả thi từ tập tin excel ta bấm “Select Files”. Giao diện sẽ hiển thị File Browser cho phép người dùng lựa chọn tập tin kết quả đã lưu ở một thư mục bất kỳ trên máy tính.

Sau đó, bấm “Open” và “Upload Files” để hoàn thành việc nhập dữ liệu. Hệ thống sẽ tự động tính toán các tham số của câu hỏi trong đề trắc nghiệm và năng lực của thí sinh tham gia kỳ thi đó.

• **Chức năng báo cáo**

Chức năng báo cáo “Xem báo cáo” cho phép lập các báo cáo kết quả tính toán các tham số của các câu hỏi theo mô hình đã chọn.

• **Định dạng dữ liệu đầu vào**

Dữ liệu đầu vào là kết quả thi chứng chỉ “Ứng dụng công nghệ thông tin cơ bản” tại Viện Công nghệ Thông tin Kinh tế, trường Đại học Kinh tế Quốc dân và được trình bày theo quy ước:

- Quy ước về đáp án: 1 – A, 2 – B, 3 – C, 4 – D (bỏ trống nếu không đáp án nào được chọn).

- Dòng đầu tiên là đáp án đúng của từng câu hỏi trong đề trắc nghiệm.

- Các dòng tiếp theo là các phương án lựa chọn của thí sinh.
- Cột 1 ở các dòng (trừ dòng 1) là số báo danh của thí sinh.

4.2. Phân tích, đánh giá và định cỡ đề trắc nghiệm bằng lý thuyết trắc nghiệm cổ điển

• Kết quả phân tích

Bảng 1. Kết quả phân tích đề trắc nghiệm bởi lý thuyết trắc nghiệm cổ điển

STT	Đáp án	Độ khó	Độ phân biệt	Ghi chú
1	B	0.95	-0.04	Đáp án có vấn đề
2	A	0.73	0.37	
3	D	0.63	0.48	
4	C	0.35	0.52	
5	D	0.5	0.44	
6	A	0.44	0.26	
7	D	0.26	0.3	
8	C	0.83	0.3	
9	B	0.34	0.33	
10	A	0.73	0.48	
11	C	0.12	0.07	
12	D	0.45	0.44	
13	A	0.83	0.22	
14	D	0.65	-0.04	Đáp án có vấn đề
15	B	0.63	0.37	
16	A	0.36	0.37	

• Định cỡ đề trắc nghiệm

- Độ tương quan có giá trị (-) thể hiện tỷ lệ thí sinh giỏi ứng đáp sai trong câu hỏi được xét nhiều hơn so với thí sinh kém, đồng nghĩa với việc có thể câu hỏi này đã có sự nhầm lẫn, cần được xem xét lại. Trong đề có 3 câu hỏi có độ tương quan (-) là câu 1, câu 14 và câu 31.

- Hệ thống phân loại câu hỏi theo 3 mức độ: khó, trung bình và dễ. Những câu được đánh giá khó (có độ khó < 0.25) có màu vàng, những câu được đánh giá trung bình (có độ khó ≥ 0.25 và ≤ 0.8) có màu trắng, những câu được đánh giá là dễ (có độ khó > 0.8) có màu xanh. Phân loại câu hỏi trợ giúp ra đề trắc nghiệm cho kỳ thi sau theo tỉ lệ câu hỏi có mức độ khó, trung bình và dễ tùy theo yêu cầu đánh giá của kỳ thi đó.

4.3. Phân tích, đánh giá và định cỡ đề trắc nghiệm bằng lý thuyết ứng đáp câu hỏi

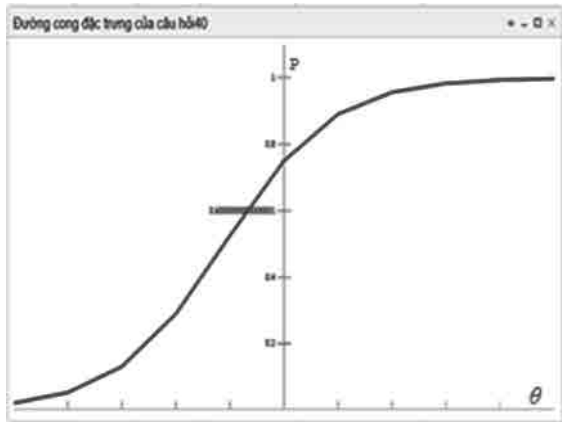
• Kết quả phân tích câu hỏi theo mô hình Rasch

Các tham số năng lực, độ khó và độ phân biệt của mô hình Rasch được tính toán dựa trên công thức 3.2. Đường cong đặc trưng câu hỏi cũng được tính dựa trên công thức 3.2 này bằng cách cho

giá trị của θ biến đổi liên tục trên trục số, ứng với giá trị θ ta tính được giá trị của $P(\theta)$.

Bảng 2. Kết quả ước lượng tham số câu hỏi theo mô hình Rasch

STT	Đáp án	Độ khó (b)	Sai số chuẩn	Đường DTCH	Ghi chú
1	B	2.47061777	0.0195668358		
2	A	-1.1055752	0.00729957735		
3	D	-1.02528393	0.00344252066		
4	C	1.115024	0.00134404306		
5	D	0.5189409	0.00233203655		
6	A	-6.35631037	0.405734529		
7	D	3.27419114	0.0296101049		
8	C	-3.93520379	0.05624743		
9	B	1.70117211	0.0133610237		
10	A	-0.8529766	0.004452264		
11	C	1.57218313	0.0025259828		
12	D	1.14548647	0.006419136		
13	A	2.74634933	0.07796219		
14	D	2.19547033	0.0662829652		
15	B	-0.6529631	0.004730427		
16	A	1.61090147	0.0101119773		



Hình 1. Đường cong đặc trưng câu hỏi (DTCH) số 40 theo mô hình Rasch

Bảng 3. Bảng ước lượng năng lực của thí sinh theo mô hình Rasch

STT	Thí sinh	Số câu đúng	Năng lực	Sai số chuẩn	Điểm thực	Điểm thực (%)	Ghi chú
1	CBC10A001	11/20	-1.3714	0.7901	9.292	0.4646	
2	CBC10A002	10/20	-0.9024	0.772	10.5009	0.525	
3	CBC10A003	13/20	-0.1895	0.7859	12.346	0.6173	
4	CBC10A004	15/20	1.6718	1.0635	15.1475	0.8074	
5	CBC10A005	6/20	-2.5032	0.9145	6.705	0.3352	
6	CBC10A006	10/20	-1.1187	0.7778	9.9382	0.4969	
7	CBC10A007	13/20	-0.1654	0.7873	12.4066	0.6203	
8	CBC10A008	15/20	1.1093	0.9434	15.2131	0.7607	
9	CBC10A009	15/20	0.8059	0.8916	14.6286	0.7314	
10	CBC10A010	12/20	-0.4676	0.7744	11.6351	0.5818	
11	CBC10A011	18/20	2.1209	1.1816	16.7616	0.8381	
12	CBC10A012	12/20	-0.3743	0.7774	11.8757	0.5938	
13	CBC10A013	14/20	-0.1295	0.7894	12.4966	0.6248	
14	CBC10A014	14/20	0.0443	0.8017	12.925	0.6463	
15	CBC10A015	9/20	-1.5840999	0.805	8.7823	0.4381	
16	CBC10A016	9/20	-1.6098	0.8071	8.69930000	0.435	

Trong đó:

- Độ nghiêng đường cong đặc trưng phản ánh độ phân biệt câu hỏi, độ nghiêng càng dốc độ phân biệt câu hỏi càng cao, ngược lại câu hỏi có độ phân biệt kém. Ý nghĩa của độ phân biệt là đối với câu hỏi cho độ dốc lớn một sự khác biệt nhỏ về năng lực thí sinh cũng gây ra một độ chênh lớn về xác suất trả lời đúng. Những câu hỏi có độ phân biệt nhỏ cần được xem xét về mặt nội dung và sẽ bị loại khỏi ngân hàng câu hỏi.

- Năng lực: Tham số năng lực θ được chuẩn hoá về thang đo chuẩn (có gốc là 0 và đơn vị là 1) sau mỗi bước lặp, các giá trị năng lực biến đổi trong khoảng từ -3 đến +3.

- Sai số chuẩn: Sai số của năng lực ước lượng năng lực, chẳng hạn, với thí sinh có mã CBC10A016, giá trị năng lực ước lượng được là -1.6098 và sai số là 0.8071. Điều này có nghĩa là, giá trị năng lực của thí sinh này thuộc trong khoảng từ (-1.6098 - 0.8071) đến (-1.6098 + 0.8071)..

- Điểm thực: tổng xác suất trả lời đúng tất cả các câu hỏi trong đề thi được tính toán theo năng lực ước lượng được của thí sinh khi thực hiện bài trắc nghiệm. Công thức tính điểm thực (True Score) cho bài trắc nghiệm của một thí sinh có năng lực θ_j như sau:

$$\text{True Score} = \sum_{i=1}^n P_i(\theta_j)$$

Với $P_i(\theta_j)$ là xác suất trả lời đúng câu hỏi thứ i của thí sinh j .

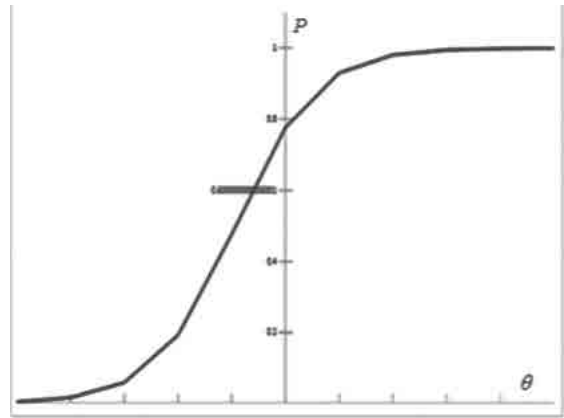
- Điểm thực (%): Tính tỷ lệ giữa phần trăm giữa điểm thực và điểm thực tối đa, nhằm quy đổi điểm thực về thang điểm 100.

- Lưu ý: Với những thí sinh có sai số chuẩn bằng 0 là những thí sinh thuộc trường hợp đặc biệt (thí sinh trả lời đúng hết tất cả câu hỏi hoặc trả lời sai tất cả câu hỏi), có ước lượng năng lực không chính xác.

• Phân tích câu hỏi theo mô hình 2 tham số

Bảng 4. Bảng kết quả ước lượng tham số câu hỏi theo mô hình 2 tham số

STT	Đáp án	Độ khó (b)	Sai số (b)	Độ phân biệt...	Sai số (a)	Đường ...	Ghi chú
1	A	-1.81740165	0.009143153	0.70787555	0.0612053424		
Nội dung		Bỏ qua	Đáp án A	Đáp án B	Đáp án C	Đáp án D	
Tỷ lệ trả lời		4	39	4	1	3	
Tương quan điểm nhị phân		0	0.36	-0.07	0	-0.14	
2	A	-0.5863521	0.01731461	0.8022886	0.05711311		
3	C	-0.332642227	0.00693008536	0.7356262	0.0228592362		
4	B	-1.56458926	0.02332094	0.422513634	0.120632343		
5	D	-1.38892233	0.0256652759	0.39026818	0.135317624		
6	C	-0.497878641	0.0115463212	0.715805769	0.03701328		
7	B	-5.329238	0.3255293	0.181750089	2.17913151		
8	C	-1.34196174	0.00501210336	0.7847292	0.0422501175		
9	D	-3.85364938	0.169194609	0.167020038	0.8022644		
10	A	-3.641822	0.147127837	0.197857678	0.778750966		
11	B	-0.568573833	0.0115654953	1.04632866	0.038710393		
12	D	-0.458004028	0.010559692	0.4432938	0.0546224043		



Hình 2. Đường cong đặc trưng của câu hỏi 40 theo mô hình 2 tham số

• Định cỡ đề trắc nghiệm

Chương trình sử dụng hai mô hình lý thuyết ứng đáp câu hỏi là mô hình một tham số **Rasch** và mô hình hai tham số để định cỡ đề trắc nghiệm, mô hình ba tham số chưa được thử nghiệm vì tham số đoán mò c chưa được xét đến trong nghiên cứu này. Trong số các kết quả ước lượng được, có một số kết quả có sai số khá lớn (cả ước lượng tham số câu hỏi và ước lượng năng lực thí sinh) do số lượng thí sinh tham gia ứng đáp câu hỏi chưa đủ lớn (Số liệu thử nghiệm được khảo sát qua thi cấp chứng chỉ Ứng dụng Công nghệ thông tin cơ bản mới được triển khai năm 2017 tại Trường Đại học Kinh tế Quốc dân).

o Về ước lượng tham số câu hỏi

Trong số 40 câu hỏi được ước lượng có một số câu hỏi cần được xem xét lại như:

- Câu hỏi số 11: có độ khó khá lớn nên xác suất trả lời đúng câu hỏi đó của thí sinh rất thấp (chỉ có 5 thí sinh có xác suất trả lời đúng > 50% và không có thí sinh nào có xác suất trả lời đúng > 80%). **Câu hỏi này cần được cân nhắc khi ra đề trắc nghiệm cho kỳ thi sau.**

- Câu hỏi số 31: **Cần phải xem xét lại xem đáp án D có phải là đáp án chính xác không, vì độ phân biệt (tương quan điểm nhị phân của đáp án D) là âm và số lượng người trả lời đúng cũng rất thấp.** Có thể có các khả năng: hoặc B không phải là đáp án đúng nhưng khi soạn câu hỏi, người soạn đã gán nhầm đáp án đúng cho câu hỏi hoặc do diễn đạt không rõ ràng khiến nhiều thí sinh thuộc nhóm có điểm số cao nhầm lẫn. Nếu cả hai trường hợp trên đều không xảy ra thì có thể trong quá trình giảng dạy giảng viên có tạo nên sự hiểu nhầm nào liên quan đến nội dung câu hỏi 31 hay không. Phương án nhiều C có khả năng là phương án đúng vì phương án C có độ tương quan lớn (số lượng thí sinh có điểm cao chọn nhiều). **Câu này cần xem xét lại độ chính xác của nó trước khi sử dụng cho đề trắc nghiệm sau.**

Bảng 5. Bảng tham số và thống kê ứng đáp của thí sinh với câu hỏi số 31

31		D	0.8549578	0.006042759		
Nội dung	Bỏ qua	Đáp án A	Đáp án B	Đáp án C	Đáp án D	
Tỷ lệ trả lời	0	20	4	52	24	
Tương quan điểm nhị phân	0	-0.11	0.07	0.19	-0.15	

o Về ước lượng năng lực thí sinh

Năng lực thí sinh được đánh giá thông qua tham số theta (θ) ước lượng được. Các giá trị năng lực trong chương trình này được chọn thang đo có gốc là 0, đơn vị là 1 và các giá trị năng lực biến đổi từ -3 đến +3, do đó có những giá trị ở đây âm là (-).

Tuy nhiên, nhìn vào các giá trị năng lực (cột giá trị theta) ước lượng được, nói chung, khó hình dung một cách trực quan về kết quả bài trắc nghiệm. Việc chuyển đổi năng lực theta sang các giá trị điểm chuẩn (True Score) giúp biểu diễn năng lực theo thang điểm trực quan.

Từ bảng kết quả ở ước lượng năng lực thí sinh cho thấy:

- Mỗi năng lực ước lượng được có sai số khác nhau. **Sai số thể hiện chính xác của phép ước lượng với từng năng lực.**

- Trong kết quả ước lượng năng lực thí sinh thu được, nên **xem xét lại năng lực của các thí sinh có sai số lớn (nếu có).**

5. Kết luận

Bài báo đã tập trung nghiên cứu quy trình định cỡ đề trắc nghiệm dựa trên lý thuyết trắc

nghiệm cổ điển và lý thuyết ứng đáp câu hỏi. Kết quả của nghiên cứu được thể hiện bởi phần mềm phân tích, đánh giá và định cỡ đề trắc nghiệm và được thử nghiệm trên kết quả thi trắc nghiệm cấp chứng chỉ Ứng dụng Công nghệ Thông tin cơ bản, Viện Công nghệ Thông tin Kinh tế, trường Đại học Kinh tế Quốc dân. Phần mềm cho phép phân tích câu hỏi trắc nghiệm dựa trên các tham số đặc trưng của lý thuyết trắc nghiệm cổ điển và đặc biệt là dựa trên mô hình một tham số và hai tham số của lý thuyết ứng đáp câu hỏi. Từ đó, đưa ra đánh giá câu hỏi có đủ khó, đủ phân biệt và có đánh giá được năng lực của thí sinh hay không. Đánh giá này làm cơ sở định cỡ đề trắc nghiệm cho các lần thi sau. Kết quả của nghiên cứu sẽ là tiền đề để xây dựng phần mềm cho phép vừa tổ chức thi trắc nghiệm, kết quả thi trắc nghiệm không chỉ đánh giá được năng lực của thí sinh mà còn là cơ sở để hình thành đề trắc nghiệm (định cỡ đề trắc nghiệm) cho lần thi kế tiếp. Trong thời gian tới tác giả sẽ tiếp tục nghiên cứu thêm các lý thuyết trắc nghiệm thích nghi (Computer Adaptive Test), lý thuyết ứng đáp câu hỏi nhiều chiều (Multiple Item Response Theory) để nâng cao hơn nữa giá trị đề trắc nghiệm.

Tài liệu tham khảo

- [1]. Dương Thiệu Tống. *Trắc nghiệm và đo lường thành quả học tập (phương pháp thực hành)*, NXB Khoa học Xã hội, 2005.
- [2]. Lâm Quang Thiệp. *Trắc nghiệm và ứng dụng*, NXB Khoa học và Kỹ thuật Hà Nội, 2008.
- [3]. Lâm Quang Thiệp. *Đo lường trong giáo dục lý thuyết và ứng dụng*, NXB Đại học Quốc gia Hà Nội, 2010.
- [4]. Nguyễn Quang Đông, Nguyễn Thị Minh. *Giáo trình kinh tế lượng*, NXB Đại học Kinh tế Quốc dân, 2012.
- [5]. Nguyễn Cao Văn, Trần Thái Ninh, Ngô Văn Thứ. *Giáo trình Lý thuyết xác suất và thống kê toán*, NXB Đại học Kinh tế Quốc dân, 2012.
- [6]. Phạm Minh Hoàn, Đoàn Minh Khoa, Trịnh Thị Linh. Tổng quan về trắc nghiệm đánh giá tham số năng lực người học, mô hình một tham số về ứng đáp câu hỏi. *Hội thảo Quốc gia Đào tạo, nghiên cứu và ứng dụng công nghệ thông tin trong phát triển kinh tế - xã hội*, 2016, tr. 49-64.
- [7]. Barker, F.B. *Item Response Theory - Parameter Estimation Techniques*, Marcel Dekker, Inc, 1992.
- [8]. Birnbaum, A., Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In: Lord, F.M. and Novick, M.R., Eds., *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading, 1968, pp. 397-479.

DEFINITION AND CORRECTION OF MULTIPLE-CHOICE QUESTIONS (QUIZ) BASED ON CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY**Abstract:**

This paper presents the basic concepts of Classical Test Theory (CTT) and Item Response Theory (IRT). Based on these fundamentals, the author presents a process of defining and correcting multiple-choice questions based on these two theories, which is the basis for correcting the value of the bank's multiple-choice questions. Research results are tested on software as a WebSite in C#. This software will analysis and evaluate multiple-choice questions to assist in correction of quiz and to correct the question banks. This study wished to continue to refine the software as a full support tool for creating the question banks, defining and correcting the multiple-choice questions.

Keywords: *Classical Test Theory, Item Response Theory, the question banks, Multiple-choice questions, quiz, Items Characteristic Function - ICF.*