

ĐO LƯỜNG-ĐÁNH GIÁ TRONG THI TRẮC NGHIỆM KHÁCH QUAN: ĐỘ KHÓ CÂU HỎI VÀ KHẢ NĂNG CỦA THÍ SINH

Nguyễn Thị Hồng Minh, Nguyễn Đức Thiện

Khoa Sau Đại học, Đại học Quốc gia Hà Nội

1. Giới thiệu

Phương pháp thi trắc nghiệm khách quan là phương pháp có nhiều ưu điểm trong việc đánh giá các kết quả của giáo dục. Người ta có thể áp dụng phương pháp này để tổ chức thi ở các lớp học thông thường với số ít thí sinh; cũng có thể áp dụng ở các kì thi quy mô lớn với hàng ngàn, hàng vạn thí sinh. Trong trường hợp qui mô lớn, bất kì một sơ suất nhỏ nào cũng có thể dẫn đến những hậu quả xấu và phản ứng xã hội bất lợi. Do đó để triển khai một kì thi quy mô lớn bằng phương pháp trắc nghiệm cần có những chuẩn bị thật kỹ lưỡng và khoa học về tất cả các khâu liên quan như đề thi, cách chấm điểm, cách đánh giá... Những kì thi như vậy được gọi là kì thi *tiêu chuẩn hoá* [5 tr.18]. Đặc biệt, các câu hỏi trong một đề thi tiêu chuẩn hoá như vậy là những câu hỏi phải đã được thử nghiệm, phân tích và *định cỡ* (tức là xác định được độ khó, độ phân biệt của từng câu và độ tin cậy, độ giá trị của cả bài). Hơn nữa, thông qua việc phân tích thống kê toàn bộ bài làm của thí sinh ta có thể thu được những số liệu quan trọng để định lượng tình hình giáo dục, đánh giá xu thế phát triển của giáo dục theo thời gian.

Bài báo trình bày tóm lược về phương pháp thi trắc nghiệm ở mục 2. Mục 3 trình bày về mô hình Rasch, một mô hình ứng đáp câu hỏi được sử dụng rộng rãi trong đánh giá giáo dục. Sử dụng mô hình Rasch cho phép tạo ra cơ sở của các phép đo lường trong giáo dục trên cả hai đối tượng là đề thi và thí sinh. Một triển khai cụ thể của phương pháp đo lường đánh giá trên cơ sở mô hình Rasch được chúng tôi trình bày trong mục 4, đó là phương pháp PROX (*approximate*), phương pháp phù hợp với những tính toán tự động bằng chương trình trên máy tính.

2. Phương pháp thi trắc nghiệm

Phần này giới thiệu phương pháp thi trắc nghiệm, ưu điểm của phương pháp thi trắc nghiệm khách quan trong đánh giá kết quả học tập, tiếp theo là các vấn đề trong việc đo lường, đánh giá đề thi trắc nghiệm bao gồm phân tích câu hỏi, phân tích bài thi.

2.1. Câu hỏi trắc nghiệm

Một cách tổng quát, các bài kiểm tra hay bài thi được chia làm hai loại: *Trắc nghiệm tự luận* và *Trắc nghiệm khách quan*. Loại trắc nghiệm tự luận đòi hỏi thí

sinh phải đưa ra câu trả lời. Trắc nghiệm khách quan đòi hỏi thí sinh lựa chọn trong số các phương án cho sẵn. Mỗi câu hỏi đưa ra chỉ có duy nhất một phương án đúng, do đó việc chấm bài sẽ không phụ thuộc vào yếu tố chủ quan của người chấm, kết quả đánh giá là chính xác và khách quan. Trắc nghiệm khách quan thường bao gồm ba loại sau:

- Loại đúng sai.
- Loại ghép đôi (hay xứng hợp).
- Loại câu hỏi có nhiều phương án trả lời để chọn.

2.2. Đo lường - Đánh giá đề thi trắc nghiệm

Việc đo lường - đánh giá đề thi trắc nghiệm đưa ra những số liệu đánh giá về các câu hỏi và toàn bài thi. Những đánh giá mang tính định lượng này sẽ hỗ trợ tốt trong việc xây dựng các ngân hàng câu hỏi.

2.2.1. Phân tích câu hỏi trắc nghiệm

Việc phân tích câu trắc nghiệm nhằm xác định độ khó, độ phân biệt hay khả năng phân loại thí sinh trả lời câu hỏi. Các giá trị này xác định được qua phân tích, tính toán trên kết quả các bài thi kiểm thử và sau đó sẽ được lưu lại như một thuộc tính của câu hỏi.

• Độ khó (DF)

Tỷ lệ thí sinh trả lời đúng trên tổng số thí sinh cho ta số đo tương đối về độ khó của câu hỏi.

Công thức tính độ khó của câu hỏi:

$$DF = \frac{R}{N} \cdot 100\% \quad (1)$$

Trong đó: R là số thí sinh trả lời đúng câu hỏi; N là tổng số thí sinh trả lời câu hỏi.

Thang phân loại độ khó quy ước như sau:

Câu dễ: DF=70-100% (tức là hầu hết các thí sinh đều trả lời đúng).

Câu trung bình: DF=30-70%.

Câu khó: DF= 0-30% (hầu hết các thí sinh đều trả lời sai).

Nên dùng các câu có DF nằm trong khoảng 25% <DF<75% (Theo [1]).

• Độ phân biệt (DI)

Phân bố tỷ lệ thí sinh trả lời đúng của các thí sinh thuộc nhóm khá, nhóm trung bình và nhóm kém cho ta số đo tương đối về độ phân biệt của câu hỏi.

Công thức tính độ phân biệt của câu hỏi:

$$DI = \frac{R_H - R_L}{N} \cdot 100\% \quad (2)$$

Trong đó: R_H là số thí sinh ở nhóm khá trả lời đúng câu hỏi; R_L là số thí sinh ở nhóm kém trả lời đúng câu hỏi; N là tổng số thí sinh trả lời câu hỏi.

Độ phân biệt $DI \leq 0$ thể hiện câu hỏi không phân biệt được thí sinh nhóm khá và kém. Theo Ebel (1956) thì những câu hỏi có $DI > 0.3$ đối với các bài trắc nghiệm trong lớp học là các câu hỏi có độ phân biệt tốt.

Công thức (2) có thể sử dụng để tính độ phân biệt cho từng phương án của mỗi câu hỏi.

2.2.2. Một ví dụ về phân tích câu hỏi

Bảng 1.3. Phân tích các câu hỏi trắc nghiệm

STT câu hỏi	N=56				Bỏ		A		B		C		D	
	R_H	R_L	DF	DI	H_0	L_0	H_A	L_A	H_B	L_B	H_C	L_C	H_D	L_D
1	14	6	36	14			4	7	14	6	7	13	3	3
2	3	1	7	4			10	14	3	1	7	4	8	9
3	10	2	21	14			6	10	8	10	4	6	10	2
4	12	6	32	11			3	6	8	11	5	5	12	6
5	21	15	64	11	1		3	3	21	15	3	8	1	2
6	0	7	13	-13			8	6	0	7	20	15	0	0
7	14	7	38	13			14	7	4	5	7	14	3	2
8	5	3	14	4			8	10	2	5	13	10	5	3
9	16	14	54	4			7	4	3	5	16	14	2	5
10	8	8	29	0			8	5	8	8	8	10	4	5

Trong đó, H_A, H_B, H_C, H_D, H_0 là số thí sinh nhóm cao (*High*) trả lời phương án A, B, C, D hoặc bỏ không làm. L_A, L_B, L_C, L_D, L_0 là số thí sinh nhóm thấp (*Low*) trả lời phương án A, B, C, D hoặc bỏ không làm.

Ở phương án có cột H và L trùng với cột R_H, R_L thì đó là phương án đúng.

Từ đó ta thu được kết quả các phương án cần phải được xem xét lại:

STT	Câu	A	B	C	D
1					X
2				X	
3					
4				X	
5		X			X
6	X				
7					X
8				X	
9		X			
10		X	X		

Trong đó, dấu “X” là các câu hỏi hoặc phương án cần phải xem xét lại.

Các phương án	Lý do cần xem xét
- 5D, 7D	- Phương án có quá ít thí sinh chọn trả lời.
- 1D, 4C, 5A, 10B	- Phương án có độ phân biệt bằng 0 ($DI=0$).
- 2C, 8C, 9A, 10A	- Phương án có độ phân biệt âm ($DI<0$).
- Câu 6	- Câu có độ phân biệt âm ($DI<0$).
- Câu 10	- Câu có độ phân biệt bằng không ($DI=0$).

Sau đó nhóm các câu hỏi, phương án có độ khó và độ phân biệt chấp nhận được sẽ được lấy làm thuộc tính của câu hỏi trong ngân hàng câu hỏi. Đây là cơ sở hỗ trợ cho việc sinh các bộ đề thi dùng cho các đối tượng thi khác nhau hoặc mục đích đánh giá khác nhau.

2.2.3. Phân tích bài trắc nghiệm

Nếu phân tích câu trắc nghiệm giúp chúng ta sửa chữa các phương án “nhiều”, làm thay đổi độ phân biệt thì phân tích bài trắc nghiệm sẽ giúp ta thay đổi độ khó của bài trắc nghiệm qua việc bổ sung, thay đổi câu hỏi.

• Độ tin cậy của bài trắc nghiệm

Là hệ số tương quan của tỷ lệ trả lời đúng-sai giữa các lần trắc nghiệm bằng các đề trắc nghiệm tương đương [1].

Tương quan giữa hai tập số X và Y được tính theo công thức sau:

Hệ số tương quan:

$$r = \frac{\overline{XY} - \overline{X}\overline{Y}}{\sqrt{[X^2 - (\overline{X})^2][Y^2 - (\overline{Y})^2]}} \quad (3)$$

Phỏng định độ tin cậy trong giáo dục được tính theo công thức ([1]):

$$r = \frac{L}{L-1} \left(\frac{1 - \sum pq}{\delta^2} \right) \quad (4)$$

Trong đó: L = Tổng số câu hỏi, p là tỷ lệ trả lời đúng câu hỏi, $q=1-p$, δ^2 là phương sai của bài thi.

• Độ giá trị của bài trắc nghiệm

Là số đo mức độ mà một bài trắc nghiệm đo được đúng mục đích mà nó định đo.

Một trong những phương pháp xác định độ giá trị là tìm mức độ tương quan giữa các bài trắc nghiệm (tiêu chuẩn hoá) khác nhau có cùng mục tiêu.

3. Mô hình RASCH

Phần này diễn giải cụ thể mô hình RASCH và ứng dụng để định cỡ câu hỏi và đo lường khả năng của thí sinh, làm tiền đề cho phương pháp PROX sẽ được trình bày trong mục 4.

3.1. Lí thuyết ứng đáp câu hỏi

Trong bất kì lĩnh vực khoa học kỹ thuật nào, để thực hiện một phép đo cũng cần tác động một thước đo lên đối tượng đo, từ đó rút ra những số đo đặc trưng cho đối tượng đó. Trước khi thực hiện một phép đo, người ta phải *thiết kế thước đo*. Động tác "*khắc độ lên thước đo*" được gọi là *định cỡ*. Để định cỡ, người ta thường cho thước đo tiếp xúc với một đối tượng đo mà đã biết tính chất. Nhằm đảm bảo độ chính xác của một phép đo lường bất kì, người ta thường nêu ra các yêu cầu chung:

- ✓ Việc định cỡ thước đo không phụ thuộc vào đối tượng đo.
- ✓ Thuộc tính đo được của đối tượng đo không phụ thuộc vào thước đo. Hiển nhiên yêu cầu về tính độc lập này cũng không phải là tuyệt đối.

Khi xét một phép đo lường cụ thể trong giáo dục:

- ✓ Thước đo có thể là một câu hỏi hoặc một bài trắc nghiệm gồm nhiều câu hỏi.
- ✓ Đối tượng đo có thể là một thuộc tính nào đó của thí sinh, chẳng hạn như năng lực về một lĩnh vực nào đó.

Thurstone (1979) cũng nêu ra những đòi hỏi của một phép đo trắc nghiệm là: "Việc định cỡ các câu hỏi không phụ thuộc vào nhóm thí sinh trả lời và số đo về thuộc tính

của một thí sinh nào đó cũng không phụ thuộc vào việc họ trả lời các câu hỏi nào” [10]. Đó là yêu cầu về tính khách quan của phép đo.

Trong thế kỷ 20 nhiều nhà nghiên cứu về đo lường trong giáo dục đã nghiên cứu mô hình hoá việc đáp ứng câu hỏi để có thể tính toán định lượng quá trình này. Lý thuyết ứng đáp câu hỏi được xây dựng trên khoa học về xác suất và thống kê với 3 giả định sau đây [5]:

- ✓ Chỉ một thuộc tính hoặc một năng lực duy nhất được đo bởi các câu hỏi trong bài trắc nghiệm. Các mô hình ứng đáp câu hỏi đo nhiều thuộc tính chưa được phát triển nhiều.
- ✓ Tính độc lập địa phương, nghĩa là năng lực được giữ không đổi trong khi làm bài trắc nghiệm, việc trả lời của thí sinh đối với hai câu hỏi nào đó là thì độc lập nhau về mặt thống kê.
- ✓ Giả định thứ ba là có một hàm đặc trưng của câu hỏi (item characteristic function- ICF) phản ánh mối quan hệ thực giữa các biến không quan sát được (năng lực) và các biến quan sát được (kết quả trả lời câu hỏi).

3.2. Mô hình RASCH

3.2.1. Giả thiết của RASCH về việc ứng đáp các câu hỏi

Khảo sát một mối quan hệ mà trong đó thí sinh thứ v có mức khả năng β_v nào đó về lĩnh vực được đo, đứng trước câu hỏi thứ t có độ khó δ_t .

Nhà Toán học người Đan Mạch, George Rasch đã đưa ra một mô hình “*ứng đáp câu hỏi*” và dùng mô hình đó để phân tích các giá trị độ đo của bài trắc nghiệm.

Rasch nói: “Một thí sinh có khả năng hơn thí sinh khác là phải có một xác suất lớn hơn để trả lời đúng một câu hỏi bất kì. Tương tự, một câu hỏi A nào đó khó hơn một câu hỏi B nghĩa là đối với bất kì thí sinh nào thì xác suất trả lời đúng câu hỏi B là lớn hơn câu hỏi A” [10].

Dựa trên giả thiết đó, Rasch đã xây dựng đường cong ứng đáp câu hỏi tương ứng. Để biểu diễn các mối quan hệ này, Rasch đã sử dụng các thủ thuật để có thể biểu diễn được khả năng của một thí sinh và độ khó của câu hỏi trên cùng một thang đo: Biểu diễn khả năng dưới dạng tỉ đối và sử dụng hàm logarit (\ln) để biến thương số thành hiệu số.

3.2.2. Hàm ứng đáp câu hỏi của Rasch

Cùng một thí sinh, có thể dễ dàng trả lời đúng câu hỏi này, nhưng lại rất khó khăn mới trả lời đúng câu hỏi khác. Rõ ràng là có một đại lượng nào đó để đặc trưng cho câu hỏi. Ta gọi là độ khó và kí hiệu là D .

Tương tự, cùng một câu hỏi, có thể dễ dàng được trả lời đúng bởi một thí sinh này, nhưng lại rất khó khăn mới trả lời đúng đối với thí sinh khác. Rõ ràng là có một đại lượng nào đó để đặc trưng cho thí sinh. Ta gọi là năng lực và kí hiệu là B .

Giả thiết B và D là 2 đại lượng liên tục. Kí hiệu $\ln B = \beta$, Kí hiệu $\ln D = \delta$.

Khi B rất lớn so với D thì $\ln \frac{B}{D} \rightarrow +\infty$

Ngược lại, khi B rất nhỏ so với D thì $\ln \frac{B}{D} \rightarrow -\infty$. Ta có: $\ln \frac{B}{D} = \beta - \delta$

Gọi P là xác suất để thí sinh có năng lực B trả lời đúng câu hỏi có độ khó D .

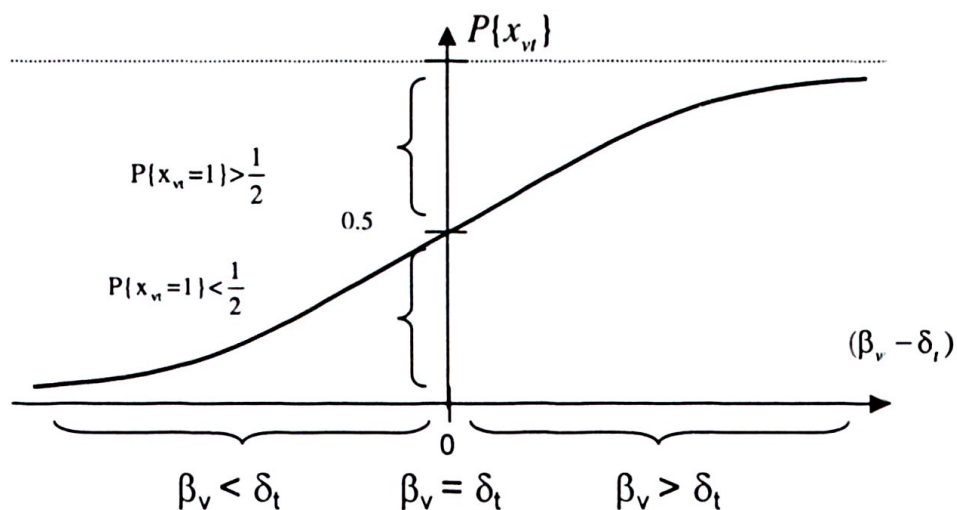
Khi $(\beta - \delta) \rightarrow +\infty$ thì $P \rightarrow 1$

Khi $(\beta - \delta) \rightarrow -\infty$ thì $P \rightarrow 0$

Khi $(\beta - \delta) \rightarrow 0$ thì $P \rightarrow 0.5$

Xác suất của thí sinh thứ v trả lời đúng câu hỏi thứ t phụ thuộc vào độ chênh giữa năng lực β_v của thí sinh và độ khó δ_t của câu hỏi, tức là hiệu số $(\beta_v - \delta_t)$. Khi $(\beta_v - \delta_t) = 0$

Đồ thị đường cong ứng đáp câu hỏi mô tả sự phụ thuộc của xác suất để thí sinh v trả lời đúng câu hỏi t với hiệu số $(\beta_v - \delta_t)$ được biểu diễn như sau:



Hình 1. Đường cong ứng đáp câu hỏi

Nhiều nhà Toán học đã đưa ra các công thức khác nhau để mô tả đường cong trên. Nhưng công thức của Rasch có ý nghĩa thực tiễn nên được biết đến nhiều nhất:

$$P\{x_{vt} = 1 | \beta_v, \delta_t\} = \frac{e^{(\beta_v - \delta_t)}}{1 + e^{(\beta_v - \delta_t)}} \quad (5)$$

Mô hình Rasch được nghiên cứu mở rộng và áp dụng rộng rãi trong khoa học xã hội nói chung và trong khoa học giáo dục nói riêng.

3.3. Sử dụng mô hình Rasch để đo lường khả năng thí sinh và định cỡ câu hỏi

Tổng quát hóa mô hình Rasch theo (5) ta có công thức xác định xác suất để thí sinh thứ v có năng lực β_v trả lời câu hỏi thứ t có độ khó δ_t với kết quả x_{vt} sẽ là:

$$P\{x_{vt} | \beta_v, \delta_t\} = \frac{e^{x_{vt}(\beta_v - \delta_t)}}{1 + e^{(\beta_v - \delta_t)}} \quad (6)$$

Trong đó x_{vt} nhận một trong hai giá trị: $x_{vt} = 1$ khi câu trả lời là đúng và $x_{vt} = 0$ khi câu trả lời là sai.

Cụ thể:

Cho một câu trả lời đúng:
$$P\{x_{vt} = 1 | \beta_v, \delta_t\} = \frac{e^{x_{vt}(\beta_v - \delta_t)}}{1 + e^{(\beta_v - \delta_t)}} \quad (7)$$

và cho câu trả lời sai:
$$P\{x_{vt} = 0 | \beta_v, \delta_t\} = \frac{1}{1 + e^{(\beta_v - \delta_t)}} \quad (8)$$

Khi có mẫu số liệu trả lời của N thí sinh trên L câu hỏi, kết quả các câu trả lời (x_{vt}) khi $v = 1..N$, $t = 1..L$. Những dữ liệu này có thể được biểu thị trong một bảng hai chiều, khi đó:

Tổng điểm của mỗi thí sinh là: $r_v = \sum_t x_{vt}$ và tổng điểm của mỗi câu hỏi là: $s_t = \sum_v x_{vt}$

áp dụng (6) cho cả bài thi ta có:

$$P\{x_{vt} | \beta_v, \delta_t\} = \prod_v \prod_t \frac{e^{x_{vt}(\beta_v - \delta_t)}}{1 + e^{(\beta_v - \delta_t)}} \quad (9)$$

Chuyển những toán tử tích số liên tục thành tổng các toán tử, phương trình (9) trở thành:

$$P\{x_{vt} | \beta_v, \delta_t\} = \frac{e^{\sum_v r_v \beta_v - \sum_t s_t \delta_t}}{\prod_v \prod_t [1 + e^{(\beta_v - \delta_t)}]} \quad (10)$$

Phương trình (10) rất quan trọng bởi nó chỉ ra rằng để đánh giá những tham số β_v và δ_i thì chúng ta cần tính tổng các lê của ma trận dữ liệu (r_v) và (s_i). Hơn nữa, các tham số này có thể được đánh giá độc lập qua sự phân định của hai công thức:

$$\sum_v^N r_v \beta_v \quad \text{và} \quad \sum_i^N s_i \delta_i$$

4. Phương pháp PROX

Trong phần này chúng tôi trình bày phương pháp PROX: cơ sở lý thuyết, các bước thực hiện trong thuật toán theo phương pháp PROX, một số kết quả thử nghiệm. Chương trình thử nghiệm đã được thực hiện với một số đề thi trắc nghiệm ở một số môn khác nhau trong trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội. Để tạo giao diện thuận lợi cho các giáo viên phối hợp sử dụng chương trình chúng tôi lựa chọn môi trường lập trình *macro* trong *Microsoft Excel*.

Khác với các phần mềm đã đóng gói hoàn chỉnh như QUEST, CONQUEST, BIGSTEP, WINSTEPS, FACET... chúng tôi hy vọng các module tính toán được thực hiện thành công sẽ dễ dàng tích hợp vào các hệ thống quản lý ngân hàng đề thi, quản lý thi trắc nghiệm phù hợp với điều kiện của Việt Nam.

4.1. Cơ sở lý thuyết của phương pháp

Phương pháp PROX được chúng tôi phát triển dựa trên mô hình Rasch theo các ý tưởng mà Wright và Douglas đã đưa ra năm 1975 [10] nhằm định cỡ độ khó các câu hỏi và đo lường khả năng của các thí sinh.

4.2. Các bước triển khai tính toán theo phương pháp PROX

• Bước 1. Nhập dữ liệu

- ✓ Nhập L_0 là tổng số câu hỏi của đề thi.
- ✓ Nhập N_0 là tổng số thí sinh thi.
- ✓ Nhập kết quả làm bài. Đánh số 1 nếu trả lời đúng, 0 nếu sai hoặc không trả lời các câu hỏi.
- ✓ Tính điểm thô của thí sinh và của câu hỏi bằng tổng các câu trả lời đúng.

• Bước 2. Sửa dữ liệu

- ✓ Loại bỏ các câu hỏi quá dễ (thí sinh nào cũng trả lời đúng) hoặc quá khó (không thí sinh nào trả lời đúng) trong quá trình tính toán tiếp theo.

✓ Tương tự, loại bỏ những thí sinh quá kém (không trả lời đúng câu hỏi nào) và quá giỏi (trả lời đúng hết tất cả các câu hỏi). Ký hiệu số câu hỏi đưa vào phân tích là L , số thí sinh là N .

• Bước 3. Sơ bộ định cỡ câu hỏi

✓ Tính tỷ lệ đúng, sai của từng câu hỏi: x_i

✓ Tính trung bình: $\bar{x} = \frac{\sum_1^L x_i}{L}$ và phương sai: $U = \frac{\sum_1^L x_i^2 - L\bar{x}^2}{L-1}$

✓ Sơ bộ định cỡ câu hỏi t theo công thức: $d_t^0 = x_t - \bar{x}$ (Vì x_0 sẽ được đặt bằng 0)

• Bước 4. Sơ bộ đo lường khả năng thí sinh

✓ Tính tần số các thí sinh có trong dải điểm và tỷ lệ trả lời đúng.

✓ Tính trung bình: $\bar{Y} = \frac{\sum_1^{L-1} n_v Y_r}{N}$ và phương sai: $V = \frac{\sum_1^{L-1} n_v Y_r^2 - \sum_1^{L-1} n_v \bar{Y}^2}{N-1}$

✓ Sơ bộ đo lường khả năng thí sinh theo công thức: $b_v^0 = y_v = \ln\left(\frac{r_v}{L-r_v}\right)$

• Bước 5. Tính các thừa số mở rộng X, Y

✓ Tính các thừa số mở rộng là hệ số phân tán độ khó của câu hỏi (X) và phân tán khả năng của thí sinh (Y) theo công thức:

$$X = \sqrt{\frac{1 + \frac{U}{1.7^2}}{1 - \frac{U.V}{1.7^4}}}; \quad Y = \sqrt{\frac{1 + \frac{V}{1.7^2}}{1 - \frac{U.V}{1.7^4}}}$$

• Bước 6. Chính xác định cỡ của câu hỏi

✓ Tính: $d_t = Y.d_t^0$

✓ Đưa định cỡ đó vào trong khoảng $[-1, 1]$.

✓ Tính sai số chuẩn: $SE(d_t) = Y \sqrt{\frac{N}{s_t(N-s_t)}} \approx \frac{2.5}{\sqrt{N}}$

• Bước 7. Chính xác hoá đo lường khả năng của thí sinh

✓ Tính: $b_v = X.b_v^0$

✓ Đưa đo lường đó vào trong khoảng $[-1, 1]$.

✓ Tính sai số chuẩn: $SE(b_v) = X \cdot \sqrt{\frac{L}{r_v(L-r_v)}} \approx \frac{2.5}{\sqrt{L}}$

Kết quả thu được trong hai bước 6 và 7 là các định cỡ chính xác của các câu hỏi d , và đo lường chính xác khả năng của các thí sinh b_v .

4.3. Một số kết quả áp dụng thực nghiệm

Sau đây là các kết quả chính của việc triển khai thực hiện chương trình trên 2 số liệu mẫu điển hình:

✓ Số liệu thứ nhất được lấy từ cuốn “*Best test design*” [9, tr 28-45] đã được đánh giá trên cơ sở định cỡ bằng tay. Chúng tôi chọn số liệu này để có thể so sánh kết quả của chương trình với các kết quả đánh giá đã đưa ra trong tài liệu.

✓ Số liệu thứ hai được lấy từ kết quả thi hết học kì môn “*Sinh lý học người và động vật*” cho sinh viên các khoá năm thứ 3, khoa Sinh học, trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội do PGS.TS Trịnh Hữu Hằng biên soạn và cho thi.

4.3.1. Thực nghiệm 1

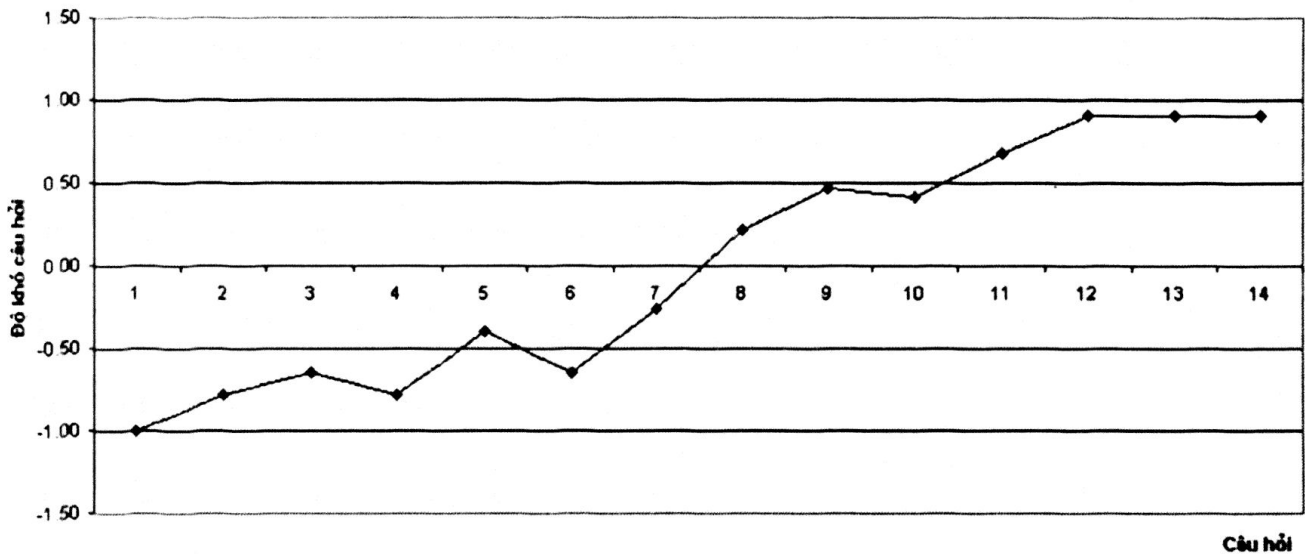
✓ Số liệu đầu vào lấy từ cuốn “*Best test design*” [10].

✓ Kích thước dữ liệu gồm 18 câu hỏi và 35 thí sinh.

✓ Quá trình xử lý : Đến bước hai, có 4 câu hỏi bị loại, trong đó 3 câu hỏi 1, 2, 3 là quá dễ (thí sinh nào cũng trả lời được) và câu hỏi 18 là quá khó (không thí sinh nào trả lời được). Có 1 thí sinh bị loại trong quá trình tính toán tiếp theo do không làm đúng câu hỏi nào (trừ 3 câu hỏi quá dễ đã bị loại). Như vậy còn có 14 câu hỏi (được đánh số lại từ 1 đến 14) và 34 thí sinh được đưa vào tính toán.

✓ Kết quả: Qua các bước xử lý 3, 4, 5 ta có được định cỡ chính xác độ khó của các câu hỏi. Đồ thị biểu diễn các số liệu định cỡ được biểu diễn trên hình 2.

✓ Nhận xét: Các kết quả tính toán được rất sát với kết quả đưa ra trong sách, trong đó ba câu hỏi 12, 13, 14 là rất khó, câu hỏi 1 là rất dễ...



Hình 2. Đồ thị biểu diễn độ khó câu hỏi với thực nghiệm 1

4.3.2. Thực nghiệm 2

✓ Kết quả thi hết học kì môn “*Sinh lý học người và động vật*” cho lớp K46, khoa Sinh học, trường Đại học Khoa học Tự nhiên.

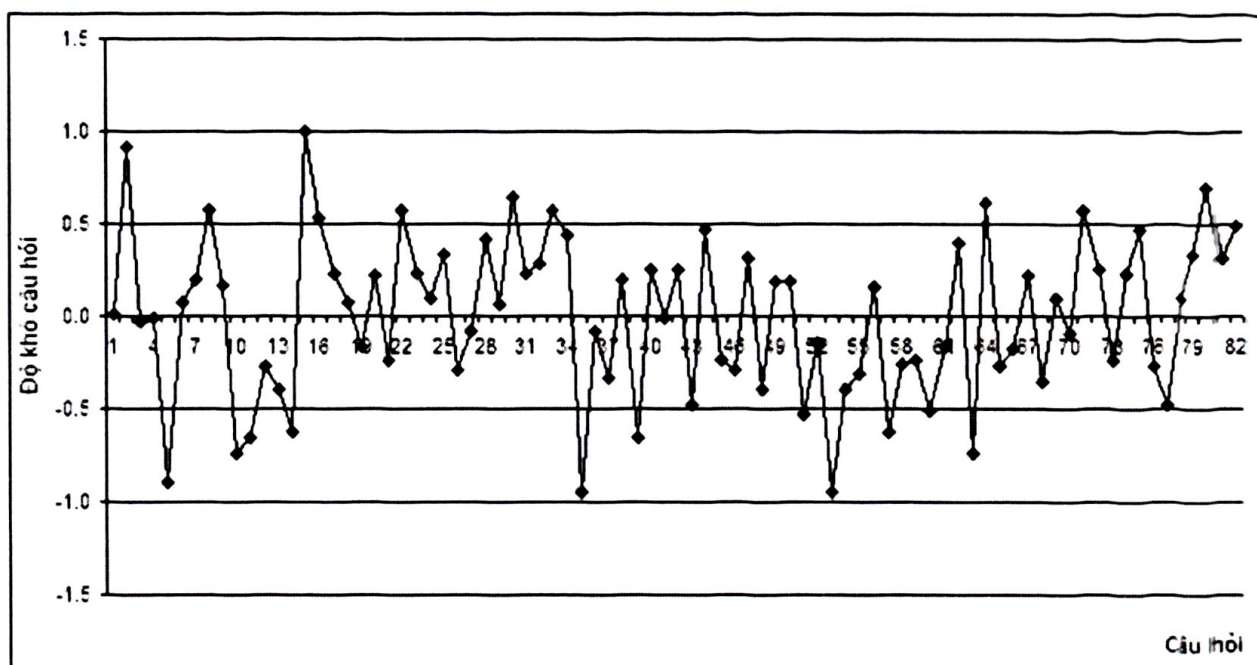
✓ Dữ liệu đầu vào gồm 90 câu hỏi và 99 thí sinh. Trong đó có 8 câu hỏi: 25, 26, 27, 36, 37, 38, 89, 90 thầy ra đề theo loại “*điền thêm*” và cho điểm lẻ nên loại ra. Như vậy còn 82 câu hỏi và 99 thí sinh đưa vào phân tích. Sau bước 2, không có thí sinh hoặc câu hỏi nào bị loại.

✓ Quá trình xử lý : Qua các bước tính toán 2,3,4,5 không có câu hỏi nào và thí sinh nào bị loại khỏi qua trình xử lý. Điều này chứng tỏ các câu hỏi trong đề thi đã được gia công tốt.

✓ Kết quả: Qua các bước xử lý 3,4,5 ta có được định cỡ chính xác độ khó của các câu hỏi. Đồ thị biểu diễn các số liệu định cỡ được biểu diễn trên hình 3.

✓ Nhận xét: Theo các kết quả tính toán độ khó cho thấy câu hỏi 15 là khó nhất, câu hỏi 35 và câu hỏi 53 là dễ nhất, tuy nhiên độ khó của các câu hỏi là chấp nhận tốt trong đề thi. Nhìn chung các câu hỏi có khả năng phân biệt (giữa các thí sinh kém và giỏi) tốt. Chúng tôi đã rút ra các nhận xét khi trao đổi với PGS.TS Trịnh Hữu Hằng về các kết quả tính toán.

✓ Ngoài 2 ví dụ nêu trên, chương trình đã được triển khai, áp dụng cho một số môn thi hết học phần như: Môn “*ngôn ngữ C*”, lớp tại chức K4, khoa Toán-Cơ-Tin học, ĐH KHTN-HN, do thầy Trần Trọng Hiếu soạn. Môn Sinh-Kỹ, Khoa tự nhiên, trường CĐSP Thái Nguyên do cô giáo Trần Thị Phúc soạn. Các nhận xét, đánh giá phù hợp với ý kiến của các giáo viên bộ môn.



Hình 3. Đồ thị biểu diễn độ khó câu hỏi trong thực nghiệm 2

5. Kết luận

Mục đích chính của phương pháp PROX là định cỡ độ khó của câu hỏi và đo lường khả năng của thí sinh. Với các câu hỏi định cỡ đã được định cỡ, người dùng (thường là các giáo viên, những người quản lý đào tạo) sẽ lấy các kết quả đó làm một thuộc tính quan trọng của câu hỏi để từ đó thành lập, bổ sung vào ngân hàng câu hỏi. Đó cũng là ưu điểm nổi bật của phương pháp, khắc phục được những hạn chế của trắc nghiệm khách quan. Việc đo lường khả năng của thí sinh giúp người quản lý đào tạo có thể đánh giá các kết quả của giáo dục và xu hướng phát triển của giáo dục theo thời gian. Hướng phát triển tiếp của đề tài là phân tích cụ thể mỗi phương án của câu hỏi và phân tích trên toàn bài thi (phân tích đề thi) làm cơ sở cho việc xây dựng chương trình quản lý ngân hàng câu hỏi và sinh đề tự động theo yêu cầu.

TÀI LIỆU THAM KHẢO

1. Lê Đức Ngọc, *Bài giảng đo lường và đánh giá thành quả học tập trong giáo dục*, NXB Đại Học Quốc Gia HN, Hà Nội, 2003.
2. Nguyễn Phụng Hoàng, Ph. D và Võ Ngọc Lan, *Phương pháp thi trắc nghiệm trong kiểm tra và đánh giá thành quả học tập*, NXB Giáo dục, Hà Nội, 1997.
3. Dương Thiệu Tống, *Trắc nghiệm và đo lường thành quả học tập, Tập 1, 2*- Bộ Giáo dục và đào tạo, Trường Đại học Tổng hợp Thành phố HCM, 1995.
4. Lâm Quang Thiệp, *Đo lường và đánh giá trong giáo dục*, khoa Sư phạm, ĐHQGHN, 2003.
5. Lâm Quang Thiệp, *Giới thiệu về đo lường và đánh giá trong giáo dục*, Bộ Giáo dục và Đào tạo, Hà Nội, 2003.

6. TS. Vũ Thị Thái, Trần Thị Hồng Chinh, Sử dụng phần mềm VIOLET xây dựng bài tập trắc nghiệm khách quan theo nội dung chương trình toán lớp 1, *Tạp chí giáo dục*, Bộ Giáo dục và Đào tạo, số 137-kì 1-5/2006, tr.19-20.
7. David Andrich- *Rasch models for measurement*- SAGE publications, 1998.
8. Travel G. Bond, Christine M. Fox- *Applying the Rasch Model*, Fundamental measurement in the Human Sciences. Lawrence Erlbaum Associates Publishers, 2001.
9. Patrick Griffin- *Mesuring Achievement Using Sub-test* from a Common Item pool. Assessment research Centre, The university of Melbourne, 1997.
10. Benjamin D. Wright, Mark H. Stone-*Best test design*-SMESA PRESSA, Chicago America, 1979, 200pp.
11. Margaret L. Wu, Raymond J. Adams, Mark R. Wilson-*ACER CONQUEST- generalised Item Response Modelling Software*- ACER Press, 1998.

VNU. JOURNAL OF SCIENCE, Nat., Sci., & Tech., T.XXII, N₀4, 2006

MEASUREMENT-ASSESSMENT IN THE OBJECTIVE TEST: QUESTION DIFFICULTY AND EXAMINEES' ABILITY

Nguyen Thi Hong Minh, Nguyen Duc Thien
School of Graduate Studies, Vietnam National University, Hanoi

The objective test is becoming popular in many subjects in different level of study. The Ministry of Education and Training has had the proposal that University entrance exams in the years to come should be in the form of the objective test. This method has many advantages in testing and measuring study results, especially in computer-based learning and testing programs (e-learning). It is, therefore, urgent and necessary to study and form theoretical basis to help scientifically measure and assess factors related to **exam questions** and **exam results**. Our article is in this study direction.

Presented in this article are our findings in experimenting PROX , a method developed from RASCH model which says: "An examinee who is more capable than another has a higher probability to answer correctly any exam question. Similarly, if a question A is more difficult than a question B, any examinee must have a higher probability to answer correctly question B than question A."

PROX method is based on the input – the results of a group of examinees who did a set of objective tests. The objectives of the method is to measure the difficulty level of exam questions and the ability of examinees. These are good parameters that will assist the administration of exam question banks and the management tests as well as the assessment of education results.

Key words: Objective test, Measurement, Ability, Difficulty