

Tính toán và so sánh độ khó của câu hỏi theo các lí thuyết khảo thí cổ điển - hiện đại bằng các phần mềm CETA/R

Vũ Đỗ Long¹, Nguyễn Văn Dũng²,
Vũ Thị Thảo³, Nguyễn Thị Mỹ Linh⁴

¹ Email: longvd@vnu.edu.vn

² Email: dungnvtkt@vnu.edu.vn

³ Email: thaovttkt@vnu.edu.vn

⁴ Email: linhtmtkt@vnu.edu.vn

Trung tâm Khảo thí Đại học Quốc gia Hà Nội
144 Xuân Thủy, Cầu Giấy, Hà Nội, Việt Nam

TÓM TẮT: Phân tích đề thi theo lí thuyết khảo thí cổ điển (CTT) và lí thuyết khảo thí hiện đại (IRT) thường đề cập đến độ khó của câu hỏi. Cách xác định độ khó câu hỏi khác nhau về biểu thức toán học nhưng có chung ý nghĩa thống kê, khi câu hỏi càng khó thì xác suất trả lời đúng câu hỏi của thí sinh càng thấp. Bài báo trình bày cách xác định độ khó cũng như các tham số khác của câu hỏi trong đề thi theo các lí thuyết nêu trên và tương quan kết quả thu được. Tính toán các giá trị tham số theo lí thuyết được lập trình trên các phần mềm Excel, CETA và R. Trên cơ sở đó, kiểm tra sự tương đồng của hai lí thuyết CTT và IRT và khuyến nghị trong thực tế.

TỪ KHÓA: Khảo thí cổ điển; khảo thí hiện đại; độ khó; độ phân biệt; hệ số đoán mò; CETA; R.
→ Nhận bài 24/8/2020 → Nhận bài đã chỉnh sửa 22/9/2020 → Duyệt đăng 05/12/2020.

1. Đặt vấn đề

Một tham số rất quan trọng khi phân tích đề thi là độ khó của câu hỏi. Tính được tham số độ khó giúp ta định lượng được câu hỏi nằm ở mức độ nào. Theo lí thuyết khảo thí cổ điển (CTT), độ khó của câu hỏi bằng tỉ lệ thí sinh làm đúng so với tổng số thí sinh; Mô hình Rash theo lí thuyết khảo thí hiện đại (IRT) cho rằng, độ khó ảnh hưởng đến xác suất trả lời đúng câu hỏi. Lí thuyết khảo thí hiện đại khẳng định năng lực thí sinh là đại lượng không đổi trong khoảng thời gian kiểm tra và được ước lượng đồng thời với độ khó (và/hoặc độ phân biệt, hệ số đoán mò) của câu hỏi thông qua xác suất trả lời đúng câu hỏi của thí sinh. Theo đó, hai phương pháp xác định độ khó, độ phân biệt của câu hỏi bằng lí thuyết khảo thí cổ điển và hiện đại là khác nhau. Ngoài ra, lí thuyết khảo thí hiện đại cũng tính được hệ số đoán mò của mỗi câu hỏi trong khi lí thuyết khảo thí cổ điển không đề cập đến. Nhóm tác giả nghiên cứu, diễn giải cách xác định các giá trị độ khó, độ phân biệt theo hai lí thuyết trên, áp dụng tính toán bằng các phần mềm Excel, Ceta và R. Sau đó, xác định tương quan giữa các kết quả thu được và đánh giá sự chính xác, mức độ khác biệt giữa các mô hình. Nghiên cứu được thực hiện với mục tiêu: 1/ Giới thiệu phương pháp và cách ước lượng các chỉ số độ khó, độ phân biệt, hệ số đoán mò của câu hỏi theo lí thuyết khảo thí cổ điển CTT và lí thuyết đáp ứng IRT; 2/ Đánh giá tương quan giữa các kết quả thu được theo hai phương pháp và các phần mềm tính toán; 3/ Khuyến nghị, đề

xuất phương pháp và phần mềm tính toán khi đánh giá câu hỏi và đề thi.

2. Nội dung nghiên cứu

2.1. Độ khó câu hỏi thi trong lí thuyết khảo thí cổ điển

Theo lí thuyết khảo thí cổ điển [1] thì độ khó (p) của một câu hỏi thi được tính như sau:

$$p = \frac{\text{số thí sinh trả lời đúng câu hỏi}}{\text{tổng số thí sinh trả lời câu hỏi}} \quad (1)$$

Theo đó, $0 \leq p \leq 1$, giá trị của p càng thấp thì câu hỏi càng khó và ngược lại. Câu hỏi rất dễ, ai cũng làm được có $p=1$; Câu hỏi rất khó, không ai làm được có $p=0$. Chẳng hạn, kì thi có 168 HS, trong đó có 82 em trả lời đúng câu hỏi 1 và 55 em trả lời đúng câu hỏi 2 thì tương ứng có độ khó câu hỏi là $p_1=82/168=0,488$; $p_2=55/168=0,327$. Câu hỏi 1 được xem như dễ hơn câu hỏi 2. Thông thường, giá trị p của câu hỏi nằm trong khoảng 0,25 - 0,75 là có thể chấp nhận được; Khi p lớn hơn 0,75 thì câu hỏi quá dễ, nhỏ hơn 0,25 thì câu hỏi quá khó.

2.2. Độ khó câu hỏi thi trong lí thuyết khảo thí hiện đại

Theo lí thuyết khảo thí hiện đại, còn được gọi là lí thuyết hồi đáp (Item Response Theory) thì độ khó của câu hỏi và năng lực của thí sinh là hai đại lượng cơ bản ảnh hưởng tới xác suất trả lời đúng câu hỏi [2].

2.2.1. Mô hình một tham số (IRT1)

Công thức trong [3]; [4]; [5] mô tả xác suất P và đại lượng θ phụ thuộc một tham số b .

$$P_i(\theta) = \frac{e^{\theta - b_i}}{1 + e^{\theta - b_i}} \quad (2)$$

Trong biểu thức này: $P_i(\theta)$ là xác suất trả lời đúng câu hỏi thứ (i) của HS có năng lực θ , $0 < P_i(\theta) < 1$; b_i là độ khó câu hỏi thứ i.

Có thể nói xác suất $P(\theta)$ phụ thuộc vào một tham số b của câu hỏi.

Nếu năng lực HS θ càng lớn thì xác suất trả lời đúng câu hỏi càng lớn, khi $\theta \rightarrow +\infty$ thì $P_i(\theta) \rightarrow 1$; khi $\theta \rightarrow -\infty$ thì $P_i(\theta) \rightarrow 0$.

Nếu độ khó câu hỏi càng lớn thì xác suất trả lời đúng câu hỏi càng nhỏ, khi $b_i \rightarrow +\infty$ thì $P_i(\theta) \rightarrow 0$; khi $b_i \rightarrow -\infty$ thì $P_i(\theta) \rightarrow 1$.

Nếu năng lực HS bằng độ khó câu hỏi thì xác suất trả lời đúng câu hỏi là 50%, khi $\theta = b_i$ thì $P = 0,5$.

Mô hình khảo thí hiện đại một tham số (thường được gọi là mô hình Rasch (1960)) khẳng định: Đường cong đặc trưng câu hỏi biểu thị phụ thuộc xác suất trả lời đúng câu hỏi P vào năng lực θ của HS, các tiệm cận ngang $P=0$; $P=1$. Độ khó câu hỏi càng lớn thì đường cong càng lệch phải. Theo mô hình một tham số các đường cong đặc trưng là đồng dạng với nhau, nói chính xác là ảnh của nhau qua một phép tịnh tiến theo phương ngang (xem Hình 1a).

2.2.2. Mô hình hai tham số

Công thức mô tả [4] có dạng

$$P_i(\theta) = \frac{e^{D a_i(\theta - b_i)}}{1 + e^{D a_i(\theta - b_i)}} \quad (3)$$

Các đại lượng $P_i(\theta)$, θ , b_i được xác định như công thức mục 2.2.1.

Hệ số $D = 1,7$ - hệ số thang đo.

Đại lượng a_i gọi là độ phân biệt câu hỏi. Cùng với sự ảnh hưởng của độ khó và năng lực của thí sinh, độ phân biệt lớn làm cho xác suất trả lời đúng của thí sinh có năng lực cao khác biệt rõ hơn (cao hơn) so với xác suất trả lời đúng của thí sinh có năng lực thấp, dẫn đến sự thay đổi của xác suất $P_i(\theta)$ càng lớn. Nếu câu hỏi có độ phân biệt

kém thì sự ảnh hưởng của độ khó câu hỏi và năng lực HS đến khả năng trả lời đúng/sai càng nhỏ. Nói cách khác, câu hỏi có độ phân biệt kém cho ta ít thông tin về thí sinh có năng lực cao hay thấp, tức là câu hỏi này không phân loại được thí sinh. Ở đây, ta hiểu xác suất $P(\theta)$ phụ thuộc hai tham số b và a của mỗi câu hỏi.

Độ phân biệt câu hỏi được xác định theo độ dốc của đường cong đặc trưng câu hỏi tại giá trị $\theta = b_i$

$$P'(\theta = b_i) = \tan \varphi = \frac{D a_i}{4}$$

Câu hỏi có độ phân biệt càng nhỏ thì đường cong đặc trưng càng có độ dốc thấp (xem Hình 1b).

2.2.3. Mô hình ba tham số

Xác suất trả lời đúng câu hỏi [4] được mô tả là:

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{D a_i(\theta - b_i)}}{1 + e^{D a_i(\theta - b_i)}} \quad (4)$$

Các đại lượng $P_i(\theta)$, θ , b_i , D , a_i được xác định như trong mục 2.2.2.

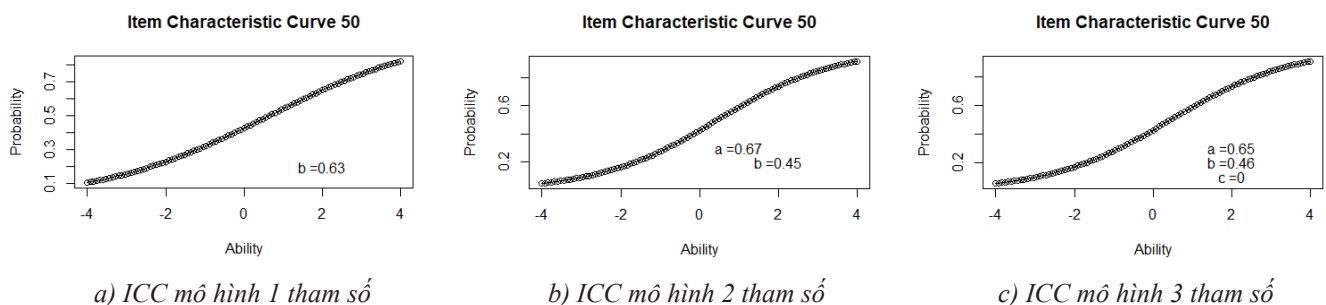
Đại lượng c_i là hệ số đoán mò câu hỏi trắc nghiệm, $0 < c_i < 1$. Khi năng lực HS thấp hoặc câu hỏi khó thì xác suất trả lời đúng câu hỏi vẫn lớn hơn c_i .

$$\lim_{\theta \rightarrow -\infty} P_i(\theta) = \lim_{b_i \rightarrow +\infty} P_i(\theta) = c_i$$

Chẳng hạn, với câu hỏi trắc nghiệm 4 phương án A,B,C,D, có 1 đáp án đúng thì dù HS có năng lực thấp hay câu hỏi rất khó đi nữa ta vẫn có xác suất trả lời đúng ngẫu nhiên/đoán mò bằng 0,25, nói cách khác trong trường hợp này thì có thể coi $c_i = 0,25$. Tương tự với câu hỏi 5 phương án trả lời - 1 đáp án đúng thì lấy $c_i = 0,20$. Đường cong đặc trưng câu hỏi thứ (i) trong mô hình 3 tham số có hai tiệm cận ngang $P=c_i$ và $P=1$. Trong mô hình này, có ba tham số của mỗi câu hỏi là b , a và c , các tham số đều ảnh hưởng đến xác suất trả lời đúng câu hỏi $P(\theta)$ (xem Hình 1c).

2.2.4. Tính toán độ khó câu hỏi mô hình một tham số theo lý thuyết khảo thí hiện đại

Dựa theo kết quả làm bài thi mà ta có thể tính trực tiếp



Hình 1: Đường cong đặc trưng của câu hỏi (ICC) theo các mô hình 1, 2 và 3 tham số

độ khó câu hỏi theo lí thuyết khảo thí cổ điển như trong 2.1. Theo lí thuyết IRT một tham số ta biết được xác suất một nhóm HS cùng điểm số/cùng mức năng lực (k) trả lời đúng một câu hỏi thứ (i) nào đó.

Theo công thức (2)

$$P_{ik} = P_i(\theta_k) = \frac{e^{\theta_k - b_i}}{1 + e^{\theta_k - b_i}} \quad (5)$$

Trong đó, $i=1,2,\dots,n$; n là số câu hỏi trong đề thi, $k=1,2,\dots,m$; m là số nhóm HS cùng mức năng lực

Từ đây ta cần xác định các giá trị độ khó câu hỏi b_i ; $\forall i = \overline{1, n}$.

Theo [6] từ biểu thức (5) ta suy ra:

$$\theta_k - b_i = \ln\left(\frac{P_{ik}}{1 - P_{ik}}\right) \quad (6)$$

Thông thường, khi số HS là lớn thì $0 < P_{ik} < 1$.

Nếu xảy ra các trường hợp đặc biệt khi $P_{ik}=0$ thì ta lấy gần đúng $P_{ik} < 1/n$; khi $P_{ik}=1$ thì ta lấy gần đúng $P_{ik} = 1 - 1/n$ để cho biểu thức (6) xác định.

Có thể giả thiết tổng khó các câu hỏi bằng không vì năng lực HS được tính theo độ khó câu hỏi.

Từ công thức (6) ta lấy tổng theo tất cả các câu hỏi thì được năng lực HS thứ (k)

$$\theta_k = \frac{1}{n} \sum_{i=1}^n \ln\left(\frac{P_{ik}}{1 - P_{ik}}\right) \quad (7)$$

Độ khó câu hỏi thứ (i) được xác định từ (6) và (7) theo mỗi nhóm năng lực (k); $\forall k = \overline{1, m}$

$$b_i = \theta_k - \ln\left(\frac{P_{ik}}{1 - P_{ik}}\right) \quad (8)$$

Vì vậy, ta lấy giá trị trung bình độ khó câu hỏi thứ (i) theo lí thuyết IRT một tham số là

$$b_i = \theta_k - \frac{1}{m} \sum_{k=1}^m \ln\left(\frac{P_{ik}}{1 - P_{ik}}\right) \quad (9)$$

Nhận xét chung: Như vậy, độ khó của câu hỏi khi tính theo các mô hình IRT1, IRT2, IRT3 có giá trị khác nhau. Theo IRT1 thì $b=0,63$, IRT2 thì $b=0,45$, IRT3 thì $b=0,46$ (xem Hình 1). Tuy sự khác biệt không lớn nhưng điều này thể hiện sự ảnh hưởng của các tham số a (độ phân biệt) và c (hệ số đoán mò) đến độ khó b của câu hỏi. Theo đó, độ phân biệt a cũng khác nhau với các mô hình hai và ba tham số, trong IRT2 thì $a=0,67$, IRT3 thì $a=0,65$.

2.3. So sánh độ khó câu hỏi theo các lí thuyết khảo thí cổ điển và hiện đại

Theo các lí thuyết cổ điển và lí thuyết hiện đại [1], ta sẽ thu được độ khó câu hỏi $x_i=p_i$ theo mục 2.1 và $y_i=b_i$ theo mục 2.2, trong đó $i=1,2,\dots,n$.

Kết quả tìm ra hai dãy X và Y là giá độ khó các câu

hỏi theo các lí thuyết khảo thí cổ điển (CTT) và lí thuyết khảo thí hiện đại (IRT).

Ta xét biểu thức r như là hệ số tương quan hai dãy X và Y thu được ở trên

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (10)$$

Miền giá trị thu được $-1 \leq r \leq 1$.

Nếu giá trị r càng gần 1 thì kết quả thu được theo các lí thuyết trên càng liên quan thuận chặt chẽ, cũng có nghĩa là câu hỏi nào khó hơn tính theo lí thuyết cổ điển cũng có độ khó cao hơn theo lí thuyết hiện đại. Trong trường hợp cần thiết, có thể lấy kết quả theo lí thuyết cổ điển dự đoán kết quả thu được theo lí thuyết hiện đại.

2.4. So sánh độ khó câu hỏi bằng phần mềm Excel

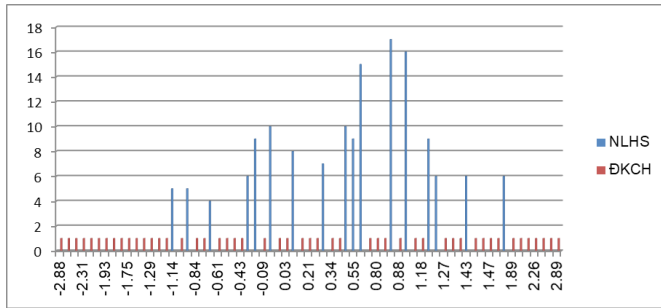
Phân tích dữ liệu bài làm của 148 sinh viên làm đề thi gồm 50 câu hỏi (n= 50) trắc nghiệm khách quan 4 lựa chọn (A, B,C, D). Theo lí thuyết khảo thí cổ điển thì độ khó câu hỏi p tính trực tiếp như xác suất HS trả lời đúng câu hỏi theo (1), độ dễ là $q=1-p$. Theo lí thuyết hiện đại một tham số thì bài toán trở nên phức tạp hơn nhiều, ta cần tìm xác suất HS trả lời đúng câu hỏi P, sau đó giải hệ phương trình thu được độ khó câu hỏi b và năng lực HS θ . Lập trình trên phần mềm Excel, dựa theo kết quả bài làm, theo lí thuyết khảo thí cổ điển CTT và IRT một tham số theo công thức (2)-(6), ta thu được độ khó câu hỏi từ câu 1 đến câu 50 trong đề thi (xem Bảng 1).

Bảng 1: Độ khó của câu hỏi tính bằng phần mềm Excel theo lí thuyết CTT và lí thuyết IRT

Câu hỏi	Độ khó/dễ của câu hỏi tính bằng phần mềm Excel		
	Độ khó theo CTT	Độ dễ theo CTT	Độ khó theo IRT
1	0.804	0.196	0,804
2	0.865	0.135	0,865
3	0.926	0.074	0,926
4	0.878	0.122	0,878
5	0.122	0.878	0,122
...
48	0.608	0.392	-1.002
49	0.804	0.196	-3.216
50	0.432	0.568	0.626

Theo CTT thì câu hỏi dễ nhất là câu số 7, có độ khó là

0,959, tương đương 95,9% HS làm đúng câu 7; câu hỏi khó nhất là câu số 5 có độ khó là 0,122, tương đương 12,2% HS làm đúng câu 5. Theo IRT một tham số thì câu hỏi dễ nhất là câu số 7, có độ khó là -2,88; câu hỏi khó nhất là câu số 5 có độ khó là 2,89. Ta thấy, tuy thang đo độ khó khác nhau nhưng CTT và IRT có sự liên hệ tương quan chặt chẽ. Với mô hình một tham số thì theo IRT ta thu được biểu đồ phân bố năng lực HS theo độ khó câu hỏi như Hình 2 bằng phần mềm Excel.

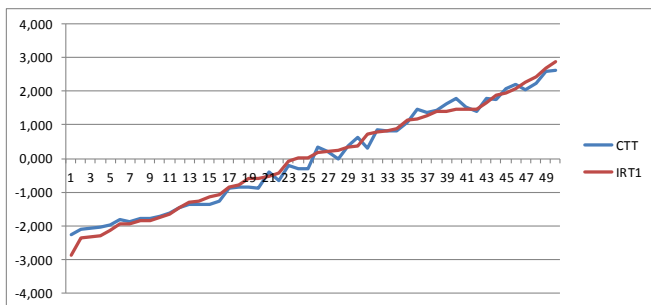


Hình 2: Biểu đồ phân bố năng lực của thí sinh và độ khó của câu hỏi

Mức năng lực HS dao động từ -1,15 đến 1,89. Biểu đồ cho thấy, có nhiều câu hỏi khó vượt năng lực các thí sinh và cũng nhiều câu hỏi dễ nằm dưới năng lực HS. HS có lượng cao nhất tập trung mức năng lực 0,86 (17 em); có đông HS đạt mức năng lực TB hơn các mức thấp và cao. Đánh giá tương quan hai đại lượng độ khó câu hỏi theo IRT (Y) và độ khó câu hỏi theo CTT (X) bởi công thức (10) có kết quả tương quan R = 0,992. Thực tế, ta thay X là độ dễ câu hỏi trong CTT, X là đại lượng là q = 1-p. Đường hồi quy tuyến tính Y = aX+b với các hệ số a và b được tính theo công thức.

$$a = \frac{\frac{1}{n} \sum x_i y_i - \bar{x} \bar{y}}{\frac{1}{n} \sum x_i^2 - \bar{x}^2} = \frac{E(XY) - EX \cdot EY}{V(X)}; b = Y - aX$$

EX = 0,429; EY = 0,000; E(XY) = 0,393; V(X) = 0,067; a = 5,838; b = -2,504



Hình 3: Biểu đồ tương quan độ khó của câu hỏi khi tính theo CTT và IRT1

Hình 3 là đồ thị trên phần mềm EXCEL biểu diễn độ khó câu hỏi, thứ tự các câu đã được hoán vị sao cho độ khó tăng dần. Đồ thị theo lý thuyết cổ điển và lý thuyết hiện đại là CTT và IRT1 tương ứng.

Biểu đồ có tung độ là độ khó câu hỏi, hoành độ là số thứ tự câu hỏi đề thi. Theo công thức (10) ta tính được giá trị r=0,992≈1 cho thấy giữa X và Y có tương quan thuận rất mạnh. Sai biến đổi đại lượng Y và aX+b có đồ thị gần như trùng nhau (xem hình vẽ). Vì vậy, có thể lấy độ khó câu hỏi theo lý thuyết khảo thí cổ điển (dễ xác định hơn) để dự đoán độ khó câu hỏi theo lý thuyết khảo thí hiện đại (khó xác định hơn).

2.5. So sánh độ khó câu hỏi bằng phần mềm CETA và R

Nhóm nghiên cứu sử dụng phần mềm CETA (Phần mềm phân tích câu hỏi, đề thi được phát triển bởi Trung tâm Khảo thí Đại học Quốc gia Hà Nội) và dựa trên lý thuyết khảo thí cổ điển [1] để phân tích câu hỏi thi và đề thi. Kết quả ước lượng độ khó thu được trong CETA trùng với kết quả tính toán trên EXCEL như Bảng 2.

Bảng 2: Độ khó của câu hỏi theo CTT tính bằng phần mềm Excel và CETA

Câu hỏi	Độ khó của câu hỏi	
	CTT - Excel	CTT - CETA
1	0.804	0,804
2	0.865	0,865
3	0.926	0,926
4	0.878	0,878
5	0.122	0,122
...
48	0.608	0,608
49	0.804	0,804
50	0.432	0,432

Trên phần mềm mã nguồn mở R (version 3.5.2), sử dụng packet *Irtoys* với gói lệnh *est* và cấu trúc *model.*PL\$est* ta tính được độ khó câu hỏi theo các mô hình (1-2-3) tham số trong mục 2.2.1 - 2.2.3. Kết quả tính toán độ khó câu hỏi trên EXCEL theo lý thuyết khảo thí hiện đại mô hình một tham số được thực hiện như trong công thức (5)-(9). Bảng kết quả thu được như sau (xem Bảng 3).

Bảng 3: Độ khó của câu hỏi tính bằng phần mềm Excel và phần mềm R theo các mô hình của lý thuyết IRT

Câu hỏi	Độ khó của câu hỏi theo IRT tính bằng phần mềm Excel và R			
	IRT1 - Excel	IRT1 - R	IRT2 - R	IRT3 - R
1	-1.281	-3.216	-1.273	-1.156
2	-1.753	-4.213	-2.077	-2.158
3	-2.336	-5.695	-2.422	-1.161
4	-1.861	-4.488	-1.381	-1.443
5	2.893	4.479	4.615	6.959
...
48	-0.089	-1.002	-1.710	0.375
49	-1.288	-3.216	-1.794	-0.808
50	0.879	0.626	0.453	0.456

(Chú thích: IRT1, IRT2, IRT3: Lần lượt là mô hình IRT 1, 2, 3 tham số)

Để kiểm tra sự chính xác của thuật toán đề xuất trong (5)-(9) ta xác định tương quan độ khó thu được theo lý thuyết khảo thí hiện đại trên EXCEL mô hình một tham số IRT1 (EXCEL) và theo lý thuyết cổ điển CTT, hiện đại IRT1, IRT2, IRT3 trên các phần mềm CETA và R (xem Bảng 4).

Bảng 4: Tương quan về độ khó của câu hỏi khi tính bằng phần mềm Excel và phần mềm R

Mô hình (phần mềm)	CTT (CETA)	IRT1 (R)	IRT2 (R)	IRT3 (R)
IRT1 (EXCEL)	r = 0,992	r = 0,995	r = 0,359	r = 0,035

(Chú thích: IRT1, IRT2, IRT3: Lần lượt là mô hình IRT 1, 2, 3 tham số)

Để xác định sự ảnh hưởng của các đại lượng khác như độ phân biệt a hay độ đoán mò c tới của độ khó câu hỏi b trong các các mô hình (1-2-3) tham số theo lý thuyết khảo thí hiện đại ta xác định tương quan độ khó thu được trên các phần mềm CETA và R theo Bảng 5.

Tài liệu tham khảo

[1] Lâm Quang Thiệp, (2010), *Đo lường trong giáo dục - lý thuyết và ứng dụng*, NXB Đại học Quốc gia Hà Nội.
 [2] Birnbaum A, (1968), *Some latent trait models and their use in inferring an examinee's ability*, Statistical theory of

Bảng 5: Tương quan về độ khó của câu hỏi khi tính bằng các phần mềm Excel, CETA, R theo các mô hình IRT

Mô hình - phần mềm	IRT1 (R)	IRT2 (R)	IRT3 (R)
CTT (CETA)	r = 0,991	r = 0,335	r = 0,021
IRT1 (R)	-	r = 0,347	r = 0,044
IRT2 (R)	-	-	r = 0,024

(Chú thích: IRT1, IRT2, IRT3: Lần lượt là mô hình IRT 1, 2, 3 tham số)

Từ số liệu trên, ta thấy sự tương quan độ khó câu hỏi giữa lý thuyết CTT với lý thuyết IRT một - hai - ba tham số (IRT1- IRT2- IRT3) tương ứng là mạnh - trung bình - yếu. Kết quả chỉ ra rằng, nếu thay đổi tham số khảo sát thì trong mô hình khảo thí hiện đại IRT thì cũng có sự thay đổi về độ khó câu hỏi.

3. Kết luận và đề xuất

Bài báo đã đề xuất cơ sở lý thuyết, công thức và các bước tính toán các chỉ số của câu hỏi như độ khó, độ phân biệt, hệ số đoán mò theo các mô hình CTT, IRT1, IRT2, IRT3. Dựa theo các biến đổi toán học với xấp xỉ trong (6) để thu được độ khó câu hỏi b_i . Kiểm nghiệm kết quả trên thực tế bằng cách tính toán b_i trên Excel theo lý thuyết CCT và lý thuyết IRT (mô hình một tham số). Với mô hình IRT hai hoặc ba tham số thì việc giải hệ phương trình với các ẩn $(\theta_k, b_i, a_i, c_i)$ gặp khó khăn khi lập trình tính toán trên Excel. Nhóm tác giả sử dụng packet **Irtoys** trên phần mềm mã nguồn mở R (version 3.5.1) để tìm ra được độ khó, độ phân biệt, hệ số đoán mò (b_i, a_i, c_i) của câu hỏi. Khi so sánh kết quả độ khó thu được ta thấy có tương quan rất mạnh giữa hai kết quả theo lý thuyết khảo thí cổ điển và hiện đại một tham số. Tuy nhiên, theo lý thuyết khảo thí hiện đại thì có sự tương quan yếu, thể hiện sự khác biệt giữa các mô hình một - hai - ba tham số.

Như vậy, lý thuyết khảo thí hiện đại đề cập đầy đủ hơn các yếu tố ảnh hưởng đến xác suất trả lời đúng câu hỏi của thí sinh như độ phân biệt, hệ số đoán mò. Tuy nhiên, việc ước lượng các giá trị này khá phức tạp, phải trải qua nhiều bước tính toán, vì vậy cần sử dụng thêm các phần mềm hỗ trợ như phần mềm Ceta, R. Trong một số trường hợp, có thể sử dụng kết quả theo lý thuyết khảo thí cổ điển như một giá trị tham khảo để so sánh với lý thuyết khảo thí hiện đại.

Mental test scores, Reading: Addison Wesley, p.395-479.
 [3] Nguyễn Thị Hồng Minh - Nguyễn Đức Thiện, (2004), *Đo lường đánh giá trong thi trắc nghiệm khách quan: Độ khó câu hỏi và khả năng của thí sinh*, Tạp chí Khoa học, Đại

- học Quốc gia Hà Nội, tr.197-214.
- [4] Ronald K. Hambleton H. Swaminathan, H. Jane Rogers, (1991), *Fundamentals of Item Response Theory*, Sage Publications.
- [5] Baker F, (2001), *The basic of item response theory*, ERIC Clearinghouse on Assessment and Evaluation.
- [6] Phạm Xuân Thanh, (2007), *Lí thuyết đánh giá (tài liệu giảng dạy lớp thạc sĩ đo lường đánh giá khóa 1, 2)*.

CALCULATION AND COMPARISON OF THE ITEM DIFFICULTY BASED ON CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY BY CETA/R PROGRAMS

**Vu Do Long¹, Nguyen Van Dung²,
Vu Thi Thao³, Nguyen Thi My Linh⁴**

¹ Email: longvd@vnu.edu.vn

² Email: dungnvtkt@vnu.edu.vn

³ Email: thaovttkt@vnu.edu.vn

⁴ Email: linhntmtkt@vnu.edu.vn

VNU - Center for Educational Testing,
Vietnam National University, Hanoi
144 Xuan Thuy, Cau Giay, Hanoi, Vietnam

ABSTRACT: *An analysis of examination papers based on Classical Test Theory (CTT) and Item Response Theory (IRT) often refers to the difficulty of the questions. The methods of determining the difficulty level by CTT and IRT are difference in mathematical expressions, but they have the same statistical meaning, that is, the more difficult the items are, the lower the examinee's probability of correctly answering the items is. This paper presents the methods to assess the difficulty and calculate the theoretical parameters of the items in the examination papers by using Excel, CETA and R programs. On that basis, the paper will define the similarity of the CTT and IRT theories, then suggest the implementation of recommendations in practice.*

KEYWORDS: *Classical Test Theory; Item Response Theory; Item difficulty; distinction; guessing coefficient; Ceta; R.*