ỨNG DỤNG MÔ HÌNH 3 THAM SỐ BIRNBAUM XÂY DỰNG HỆ THỐNG TỰ ĐỘNG ĐÁNH GIÁ NĂNG LỰC NGƯỜI HỌC

APPLICATION OF THREE-PARAMETER MODEL BIRNBAUM (3PM) TO BUILD AN AUTOMATIC SYSTEM FOR EVALUATING LEARNERS' CAPABILITY

Mai Văn Hà, Đặng Hoài Phương

Trường Đại học Bách khoa, Đại học Đà Nẵng

mvha@dut.udn.vn; dhphuong@dut.udn.vn

Tóm tắt - Hiệu quả quá trình học tập phụ thuộc khách quan vào quá trình đánh giá năng lực người học và trắc nghiệm trên máy tính là một trong những phương pháp đánh giá được sử dụng phổ biến hiện nay. Một trong những phương pháp trắc nghiệm cho phép đưa ra kết quả đánh giáchính xác, hiệu quả và khách quan là mô hình trắc nghiệm thích nghi. Các hệ thống trắc nghiệm thích nghi trên máy tính hiện nay được xây dựng dựa trên cơ sở các thuật toán trắc nghiệm thích nghi khác nhau và phổ biến hơn là thuật toán trắc nghiệm thích nghi dựa trên lý thuyết đáp ứng câu hỏi (Item Response Theory - IRT). Với giải pháp này tồn tại ba mô hình toán học trong lý thuyết đáp ứng câu hỏi và việc áp dụng mô hình 3 tham số Birnbaum (3PM): độ khó, độ phân biệt và độ dự đoán để xây dựng hệ thống trắc nghiệm thích nghi là linh hoạt nhất vì nó cho phép đánh giá năng lực người học tốt hơn với lượng thông tin tối đa.

Từ khóa - trắc nghiệm thích nghi; Lý thuyết đáp ứng câu hỏi; Thuật toán trắc nghiệm thích nghi; Hệ thống trắc nghiệm thích nghi; Mô hình 3 tham số BirnBaum.

Abstract - The efficiency of a learning process depends objectively on the process of evaluating learners' capability, and computerized testing system is one of popular methods nowadays. The most popular one is adaptive testing which can evaluate learner knowledge accurately and objectively. Nowadays, computerized adaptive testing is based on different adaptive algorithms, and the most popular one is Item Response Theory. With this solution, there are three mathematical models in the Item Response Theory and the application of three parameter model Birnbaum with levels of difficulty, discrimination and prediction to build the adaptive testing system is most flexible because it gives a better way to evaluate learners with maximum information.

Key words - adaptive testing; Item Response Theory; Algorithm of adaptive testing; Adaptive testing system; Three-parameter model (3-PM).

1. Đặt vấn đề

Hiện nay, trắc nghiệm là một trong những phương pháp đánh giá kiến thức người học đang được sử dụng rộng rãi và mang lại hiệu quả tốt, đặc biệt là các hệ thống trắc nghiệm trên máy tính. Trước đây, các hệ thống trắc nghiệm chủ yếu được xây dựng trên cơ sở lý thuyết trắc nghiệm cổ điển (Classical Test Theory - CTT).

Lý thuyết trắc nghiệm cổ điểnlà một trong những lý thuyết quan trọng liên quan đến khoa học đo lường đánh giá được xây dựng dựa trên khoa học xác suất và thống kê [1]. Lý thuyết này được ứng dụng trong việc phân tích câu hỏi trắc nghiệm nhằm làm tăng chất lượng của chúng, loại bỏ, sửa chữa và tuyển chọn câu hỏi theo yêu cầu. Tuy nhiên, việc ứng dụng lý thuyết trắc nghiệm cổ điển vào các hệ thống trắc nghiệm không đánh giá được mức độ năng lực thí sinh một cách chính xác. Tất cả các thí sinh đều phải trả lời một số lượng câu hỏi như nhau và các câu hỏi được lựa chọn ngẫu nhiên từ ngân hàng câu hỏi. Điều này sẽ khiến thí sinh sẽ không thể nhận được câu hỏi phù hợp với mức độ năng lực hiện tại của mình và dẫn đến việc đánh giá không chính xác.

Nhằm khắc phục nhược điểm của Lý thuyết trắc nghiệm cổ điển thì Trắc nghiệm thích nghi được phát triển. Trắc nghiệm thích nghi [2] là quá trình trắc nghiệm mà trong đó tập hợp câu hỏi được đưa ra để đánh giá phụ thuộc vào khả năng của thí sinh. Với việc sử dụng trắc nghiệm thích nghi thì thứ tự, số lượng, nội dung và độ khó câu hỏi đưa ra cho mỗi thí sinh là khác nhau, cho phép thí sinh có thể nhận được câu hỏi phù hợp với mức độ năng lực hiện tại của mình nhất, giúp cho kết quả đánh giá mức độ năng lực thí sinh được chính xác và khách quan hơn.

Hiện nay, các hệ thống trắc nghiệm thích nghi trên máy tính được xây dựng dựa trên các thuật toán trắc nghiệm thích nghi khác nhau. Một trong những mô hình được sử dụng phổ biến nhất hiện nay là mô hình trắc nghiệm thích nghi trên cơ sở lý thuyết đáp ứng câu hỏi (Item Response Theory – IRT) [3].

Lý thuyết đáp ứng câu hỏicho phép đánh giá mối tương quan giữa mức độ năng lực thí sinh với tập hợptham số câu hỏi thông qua hàm thông tin câu hỏi. Do đó số lượng, thứ tự và mức độ câu hỏi đánh giá đối với mỗi thí sinh là khác nhau.

Từ [1] và[3] có thể thấy rằng lý thuyết trắc nghiệm cổ điển tiếp cận việc đánh giá thí sinh ở cấp độ tập hợp các câu hỏi trong một đề thi trắc nghiệm, còn lý thuyết đáp ứng câu hỏi lại tiếp cận ở cấp độ từng câu hỏi được đưa ra.Ưu điểm của trắc nghiệm thích nghi so với lý thuyết trắc nghiệm cổ điển là mỗi thí sinh sẽ trả lời một bộ câu hỏi riêng biệt, khác nhau về nội dung, độ khó và số lượng câu hỏi; và mỗi thí sinh sẽ được đánh giá riêng biệt (dựa vào mức độ năng lực của thí sinh).

Trong bài báo này, nhóm tác giả đề xuất sử dụng mô hình 3 tham số Birnbaum [4] để xây dựng hệ thống trắc nghiệm thích nghi đánh giá năng lực thí sinh trên máy tính.

2. Xây dựng mô hình trắc nghiệm thích nghi dựa trên mô hình 3 tham số Birnbaum

Lý thuyết đáp ứng câu hỏi dựa trên phân tích cấu trúc tiềm ẩn (LSA). Về bản chất, lý thuyết đáp ứng câu hỏi được xây dựng dựa trên mô hình toán học [5] sử dụng các mô hình logistic, mô tả sự phụ thuộc giữa xác suất trả lời đúng câu hỏi trắc nghiệm đưa ra với mức độ năng lực hiện tại của thí sinh.

Tồn tại 3 mô hình toán học trong lý thuyết đáp ứng câu hỏi [5]: mô hình 1 tham số (1PM), mô hình 2 tham số (2PM) và mô hình 3 tham số Birnbaum (3PM). Các mô hình trên chỉ khác nhau về số lượng tham số của câu hỏi.

Mô hình 1 tham số [6] chỉ đề cập đến sự phụ thuộc giữa xác suất trả lời đúng câu hỏi của thí sinh và độ khó câu hỏi b, mô hình 2 tham số [4] thì xác suất trả lời đúng câu hỏi phụ thuộc vào độ khó b và độ phân biệt a. Rõ ràng trong trường hợp câu hỏi trắc nghiệm nhị phân thì thí sinh có thể dự đoán để trả lời câu hỏi, vì vậy mô hình 3 tham số ra đời để khắc phục nhược điểm trên [4]. Để đánh giá chính xác hơn sự phụ thuộc của xác suất trả lời đúng câu hỏi vào tập hợp tham số của câu hỏi thì cần bổ sung tham số độ dự đoán c.

Như vậy mô hình trắc nghiệm thích nghi được xây dựng dựa trên mô hình 3 tham số Birnbaum sẽ linh hoạt và chính xác hơn các mô hình khác.

Theo mô hình 3 tham số của Birnbaum thì xác suất trả lời đúng câu hỏi i của thí sinh j có năng lực θ_j được tính như sau:

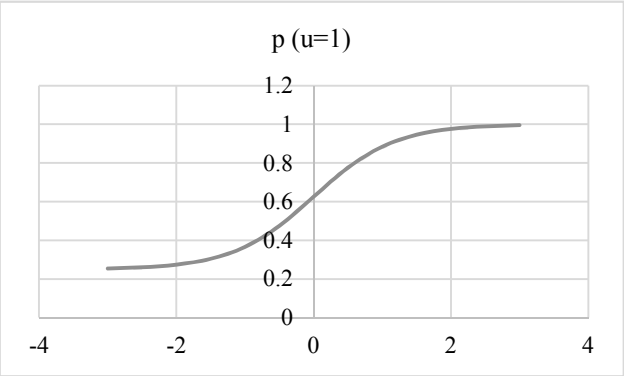
$$p(a_i, b_i, c_i) = c_i + (1 - c_i)\frac{1}{1+e^{-1.7a_i(\theta_j - b_i)}} \quad (1)$$

Trong đó:

$p(a_i, b_i, c_i)$ là xác suất trả lời đúng câu hỏi thứ i của thí sinh j;

a_i, b_i, c_i là tập hợp các tham số của câu hỏi i: độ phân biệt, độ khó và độ dự đoán của câu hỏi i;

θ_j là mức độ năng lực hiện tại của thí sinh j.



Hình 1. *Đồ thị mô tả sự phụ thuộc hàm xác suất trả lời đúng câu hỏi (a=1, b=0, c= 0.25) của người học có năng lực θ*

Với các tham số a_i, b_i, c_i đã biết của câu hỏi i có thể xây dựng đồ thị phụ thuộc giữa hàm xác suất trả lời đúng câu hỏi với năng lực của thí sinh (Hình 1).

Tiệm cận dưới của hàm xác suất trả lời đúng câu hỏitiến về giá trị độ dự đoán c của câu hỏi, có nghĩa là nếu như người học không biết trả lời câu hỏi thì xác suất lựa chọn đáp án đúng cho câu hỏi (xác suất trả lời đúng câu hỏi) bằng c.

Tham số b của câu hỏi xác định điểm uốn của đồ thị mà không làm thay đổi đường cong của nó. Tham số này xác định độ khó của câu hỏi, với giá trị b càng lớn thì xác suất trả lời đúng câu hỏi của thí sinh càng giảm và ngược lại, giá trị b càng nhỏ thì xác suất trả lời đúng câu hỏi đó càng lớn.

Tham số độ phân biệt a của câu hỏi xác định góc nghiêng của đường cong hàm xác suất trả lời đúng câu hỏi tại điểm uốn (b = θ). Nó cho thấy khả năngcâu hỏi có thể phân biệt được các thí sinh có các mức độ năng lực khác nhau, hay có thể nói độ phân biệt câu hỏi là đại lượng ảnh hưởng đến xác suất trả lời đúng câu hỏi giữa các thí sinh có mức độ năng lực khác nhau.

Trong các mô hình trắc nghiệm thích nghi ứng dụng IRT hiện nay phần lớn đều dựa trên mô hình 2 tham số (độ khó và độ phân biệt) vì sử dụng các loại câu hỏi trắc nghiệm khác nhau. Tuy nhiên để tăng thêm tính chính xác của quá trình đánh giá, nhóm tác giả đề xuất sử dụng loại câu hỏi trắc nghiệm khách quan nhiều lựa chọn. Với câu hỏi trắc nghiệm khách quan nhiều lựa chọn, mỗi câu hỏi trắc nghiệm sẽ được bổ thêm tham số c là độ dự đoán câu trả lời của câu hỏi. Áp dụng mô hình trắc nghiệm thích nghi dựatrênlý thuyết đápứngcâuhỏivới mô hình 3 tham số Birnbaumvới câu hỏi trắc nghiệm khách quan nhiều lựa chọn đượcthựchiệntheo các bước như sau:

– *Bước 1*: Thiết lập mức độ năng lực ban đầu cho thí sinh;
– *Bước 2*: Lặp lại từ bước 3 đến bước 5 cho đến khi gặp điều kiện dừng;
– *Bước 3*: Đánh giá tập hợp câu hỏi chưa đưa ra cho thí sinh và tính toán hàm thông tin của các câu hỏi đó dựa trên mức độ năng lực hiện tại của thí sinh;
– *Bước 4*: Lựa chọn và đưa ra câu hỏi phù hợp nhất với mức độ năng lực hiện tại của thí sinh;
– *Bước 5*: Đánh giá lại năng lực thí sinh trên cơ sở kết quả trả lời câu hỏi đưa ra.

Mức độ năng lực ban đầu của mỗi thí sinh là không xác định, và thường được lấy giá trị trung bình bằng 0.

Lựa chọn câu hỏi phù hợp với mức độ năng lực hiện tại của thí sinh là rất quan trọng. Nếu câu hỏi đưa ra là quá khó hay quá dễ đối với thí sinh thì sẽ mang lại ít thông tin về mức độ năng lực thí sinh. Để đánh giá được chính xác mức độ năng lực thí sinh cần thiết phải đưa ra câu hỏi với bộ tham số mang lại thông tin tối đa trong việc đánh giá mức độ năng lực thí sinh.

Theo Birnbaum, mỗi câu hỏi trắc nghiệm cung cấp một lượng thông tin nào đó về mức độ năng lực của thí sinh. Birnbaum đề xuất hàm thông tin $I_i(\theta)$ của câu hỏi i được tính toán phụ thuộc vào năng lực θ của thí sinh theo công thức sau:

$$I_i(\theta) = \frac{\left(p_i'(u_k=1|\theta,a_i,b_i,c_i)\right)^2}{p_i(u_k=1|\theta,a_i,b_i,c_i)\left(1-p_i(u_k=1|\theta,a_i,b_i,c_i)\right)} \quad (2)$$

Trong đó:

– $p_i(uk = 1|\theta, a_i, b_i, c_i)$ là xác suất trả lời đúng câu hỏi i với tập hợp tham số: a_i, b_i, c_i;
– $p_i'(uk = 1|\theta, a_i, b_i, c_i)$ là đạo hàm bậc nhất của xác suất trả lời đúng câu hỏi i theo mức độ năng lực.

Như vậy, trong bước thứ nhất, giá trị hàm thông tin $I_i(\theta)$ được tính cho tập hợp các câu hỏi chưa đưa ra cho thí sinh. Phương pháp phổ biến nhất trong việc lựa chọn câu hỏi tiếp theo phù hợp với mức độ năng lực hiện tại của thí sinh là sử dụng hàm thông tin lớn nhất.

Sau khi trả lời mỗi câu hỏi thì vấn đề đặt ra là làm sao đánh giá lại mức độ năng lực của thí sinh vì giá trị mức độ năng lực mới này sẽ được sử dụng để lựa chọn câu hỏi tiếp theo phù hợp với thí sinh. Sử dụng phương pháp Maximum likelihood [7] để đánh giá mức độ năng lực của thí sinh sau mỗi câu trả lời, tức là tìm một giá trị mức độ năng lực mà hàm sau đây là tối đa:

$$f(u_1, u_2, \dots, u_n, \theta) = p_i(u_1 = 1, \theta)^{u_1}. p_i(u_1 = 0, \theta)^{1-u_1}. p_i(u_2 = 1, \theta)^{u_2}. p_i(u_2 = 0, \theta)^{1-u_2} \dots p_i(u_n = 1, \theta)^{u_n}. p_i(u_n = 0, \theta)^{1-u_n} \to max$$

Hay nói cách khác:

$$f(u_1, u_2, \dots, u_n, \theta) = \prod_{k=1}^{n} p_i(u_k, \theta) \to max \quad (3)$$

Trong đó:

$u_1, u_2, \dots, u_n$ là tập hợp câu trả lời của thí sinh đối với các câu hỏi đã đưa ra;

$p_i(u, \theta)$ là xác suất trả lời đúng câu hỏi i của thí sinh có mức độ năng lực θ, được tính theo công thức (1).

Trắc nghiệm được bắt đầu với mức độ năng lực của thí sinh là θ_s, từ công thức (3) có thể tiến hành đánh giá lại mức độ năng lực thí sinh sau khi trả lời một câu hỏi θ_{s+1}theo công thức sau:

$$\theta_{s+1} = \theta_s + \frac{\sum_{i\epsilon n} S_i(\theta_s)}{\sum_{i\epsilon n} I_i(\theta_s)}$$

Trong đó: n là tập hợp bộ câu hỏi thí sinh đã trả lời;

$S_i(\theta_s)$ được tính theo công thức:

$$S_i(\theta_s) = \frac{(u_i - p_i(u_i = 1|\theta_s, a_i, b_i, c_i)).p_i'(u_i = 1|\theta_s, a_i, b_i, c_i)}{p_i(u_i = 1|\theta_s, a_i, b_i, c_i)(1 - p_i(u_i = 1|\theta_s, a_i, b_i, c_i))}$$

+ $p_i(u_i = 1|\theta_s, a_i, b_i, c_i)$ và $p_i'(u_i = 1|\theta_s, a_i, b_i, c_i)$ là xác suất trả lời đúng câu hỏi i và đạo hàm bậc nhất tương ứng của xác suất đó;

+ $I_i(\theta s)$ – hàm thông tin của câu hỏi i được tính theo công thức (2).

Quá trình trắc nghiệm đánh giá mức độ năng lực của mỗi thí sinh kết thúc khi đã đưa ra được chính xác mức độ năng lực của thí sinh đó. Trong quá trình trắc nghiệm thích nghi theo mô hình đề xuất, sau mỗi câu trả lời mức độ năng lực thí sinh sẽ được đánh giá lại. Nếu sự chênh lệch mức độ năng lực thí sinh giữa hai lần đánh giá liên tiếp là rất nhỏ thì có thể xem đó là kết quả đánh giá cuối cùng về mức độ năng lực thí sinh. Bởi vì nếu thí sinh có trả lời thêm các câu hỏi nữa thì mức độ năng lực thí sinh gần như không đổi và đó cũng chính là kết quả đánh giá mức độ năng lực cuối cùng của thí sinh.

Mô hình trắc nghiệm thích nghi trên sử dụng mô hình xác suất dựa trên kết quả thống kê tập hợp tham số câu hỏi trong ngân hàng câu hỏi trắc nghiệm, do đó cho phép đánh giá mức độ năng lực thí sinh được khách quan và chính xác hơn.

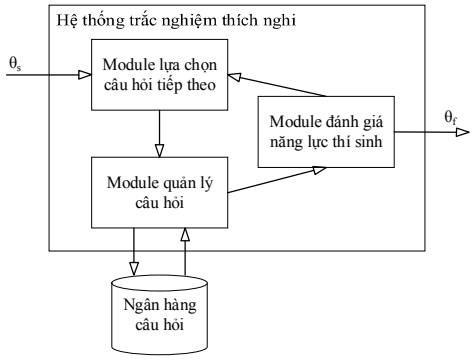
3. Xây dựng hệ thống trắc nghiệm thích nghi

Nhóm tác giả tiến hành xây dựng hệ thống trắc nghiệm thích nghi trên cơ sở mô hình đề xuất, sử dụng mô hình client-server và mô hình Model View Controller (MVC).

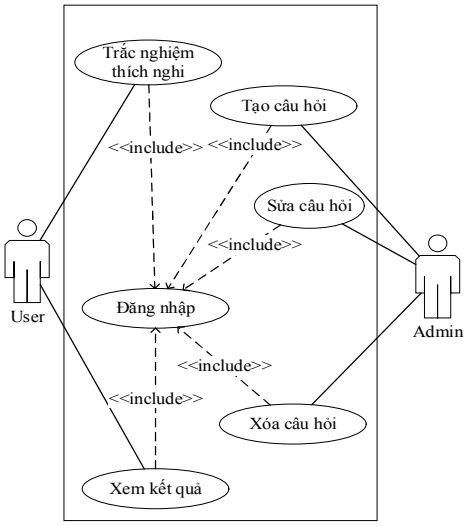
Hệ thống trắc nghiệm thích nghi gồm hai module chính: quản lý ngân hàng câu hỏivà trắc nghiệm thích nghi theo mô hình 3 tham số Birnbaum.

Kiến trúc tổng thể của hệ thống trắc nghiệm thích nghi (Hình 2).

Biểu đồ ca sử dụng của hệ thống trắc nghiệm thích nghi được mô tả như Hình 3.

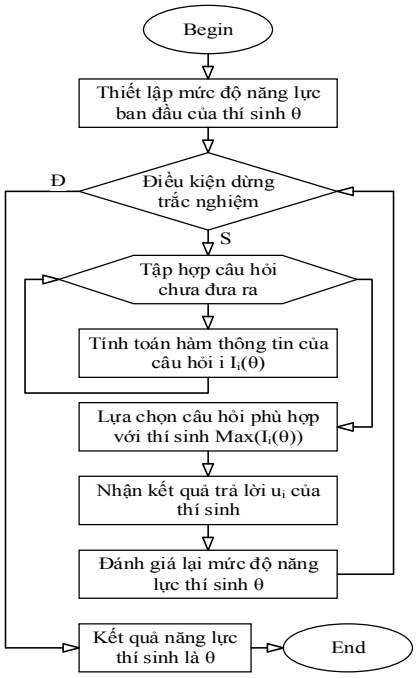


Hình 2. *Kiến trúc tổng thể hệ thống trắc nghiệm thích nghi.*



Hình 3. *Biểu đồ ca sử dụng của hệ thống trắc nghiệm thích nghi*

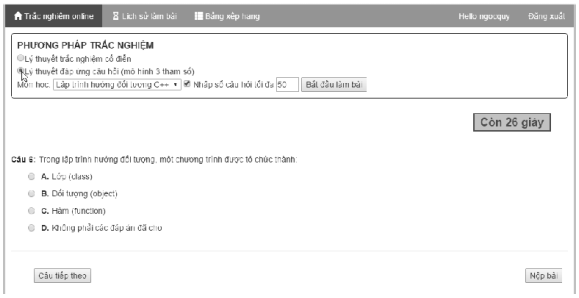
Thuật toán trắc nghiệm thích nghi trên cơ sở mô hình 3 tham số Birnbaum được mô tả như Hình 4:



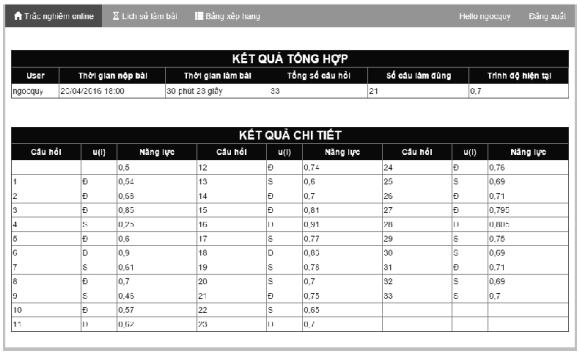
Hình 4. *Thuật toán trắc nghiệm thích nghi áp dụng mô hình 3 tham số Birnbaum trong lý thuyết đáp ứng câu hỏi.*

Nhóm tác giả đã tiến hành triển khai hệ thống trong thực tế đánh giá mức độ năng lực sinh viên với ngân hàng câu hỏi

trắc nghiệm của môn học Lập trình hướng đối tượng C++ và điều kiện dừng của hệ thống trắc nghiệm thích nghi là độ biến thiên mức độ năng lực thí sinh $\Delta\theta \leq 0{,}01$. Giao diện của hệ thống trắc nghiệm thích nghi đánh giá năng lực của một thí sinh được thể hiện ở Hình 5 và Hình 6.



Hình 5. *Giao diện hệ thống trắc nghiệm thích nghi đánh giá năng lực thí sinh*



Hình 6. *Giao diện kết quả chi tiết quá trình đánh giá năng lực thí sinh*

Kết quả quá trình trắc nghiệm đánh giá mức độ năng lực của một thí sinh được mô tả ở bảng sau:

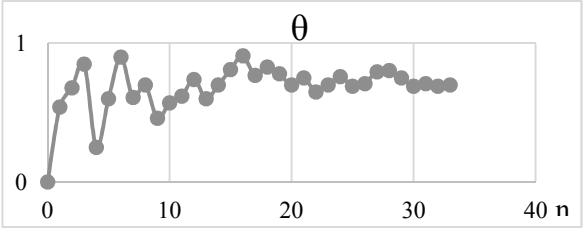
Bảng 1. *Kết quả chi tiết quá trình đánh giá năng lực một thí sinh bằng hệ thống trắc nghiệm thích nghi*

Câu hỏi	u_i	θ	Câu hỏi	u_i	θ
θ_s		0,0	17	S	0,77
1	Đ	0,54	18	Đ	0,83
2	Đ	0,68	19	S	0,78
3	Đ	0,85	20	S	0,7
4	S	0,25	21	Đ	0,75
5	Đ	0,6	22	S	0,65
6	Đ	0,9	23	Đ	0,7
7	S	0,61	24	Đ	0,76
8	Đ	0,7	25	S	0,69
9	S	0,46	26	Đ	0,71
10	Đ	0,57	27	Đ	0,795
11	Đ	0,62	28	Đ	0,805
12	Đ	0,74	29	S	0,75
13	S	0,6	30	S	0,69
14	Đ	0,7	31	Đ	0,71
15	Đ	0,81	32	S	0,69
16	Đ	0,91	33	Đ	0,7

Trong đó: $\theta_s = 0$ là mức độ năng lực ban đầu của thí sinh;

u_i là kết quả trả lời của mỗi câu hỏi, nếu u_i = Đ – trả lời đúng câu hỏi và u_i = S trong trường hợp ngược lại.

Từ Bảng 1 nhóm tác giảđã đưa ra đồ thị quá trình đánh giá năng lực thí sinh như Hình 7:



Hình 7. *Đồ thị mô tả quá trình đánh giá mức độ năng lực thí sinh trong quá trình trắc nghiệm thích nghi*

Dựa vào đồ thị thì kết quả mức độ năng lực của thí sinh rơi vào khoảng 0,7 sau 33 câu hỏi. Theo đồ thị có thể thấy ở những câu hỏi đầu tiên, độ biến thiên của mức độ năng lực thí sinh là rất lớn, nhưng từ câu hỏi 20 trở đi thì độ biến thiên của mức độ năng lực thí sinh bắt đầu giảm xuống.

Quá trình đánh giá mức độ năng lực thí sinh sẽ kết thúc khi không thể đánh giá được nữa hay nói cách khác là mức độ năng lực thí sinh sẽ gần như không thay đổi với các câu hỏi tiếp theo. Sau khi kết thúc quá trình đánh giá thì mức độ năng lực cuối cùng của thí sinh sẽ được quy đổi bằng một thang đo khác trong lý thuyết dáp ứng câu hỏi [5].

4. Kết luận

Hệ thống trắc nghiệm thích nghi trên cơ sở mô hình 3 tham số Birnbaum trong lý thuyết đáp ứng câu hỏi cho phép đánh giá một cách liên tục mức độ năng lực thí sinh trong quá trình trắc nghiệm bằng cách sử dụng hiệu quả tất cả các thông tin có sẵn về mức độ năng lực thí sinh và tập hợp tham số câu hỏi trong cơ sở dữ liệu.

Vì vậy, kết quả đánh giá năng lực thí sinh sẽ chính xác và khách quan hơn so với lý thuyết trắc nghiệm cổ điển. Thông qua việc sử dụng hệ thống trắc nghiệm thích nghi trên cơ sở mô hình đề xuất sẽ góp phần nâng cao hiệu quả của quá trình đào tạo, đặc biệt là đào tạo trực tuyến.

Tuy nhiên, hệ thống trắc nghiệm thích nghi đề xuất vẫn còn tồn tại nhược điểm là chưa đánh giá được sự ảnh hưởng chuỗi câu trả lời trước đó đến kết quả đánh giá năng lực của thí sinh. Vấn đề này sẽ được nhóm tác giả tiếp tục nghiên cứu và giải quyết.

TÀI LIỆU THAM KHẢO

[1] Lâm Quang Thiệp, Trắc nghiệm và Ứng dụng, NXB Khoa học và Kỹ thuật, 2008.

[2] J. Tian, D. Miao, X. Zhu, and J. Gong, An introduction to the computerized adaptive testing, US-China Education Review, vol. 4, no. 1, pp. 72–81, 2007.

[3] Lê Xuân Tài, Đặng Hoài Phương, Xây dựng mô hình trắc nghiệm thích nghi trên cơ sở lý thuyết đáp ứng câu hỏi, Tạp chí Khoa học Đại học Huế, T. 97, Số 9, 2015.

[4] Allan Birnbaum, Some latent trade models and their use in inferring an examinee's ability, M.A: Addison-Wesley, 1968.

[5] Lord, F.M, Applications of Item Response Theory to Practical Testing Problems. Lawrence Erbaum Associates, Publishers, 1980.

[6] Rasch, G., Probablistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Danish Institute for Educational Research, 1960.

[7] K. Wauters, P. Desmet, and W. Van Den Noortgate, "Adaptive item-based learning environments based on the item response theory: Possibilities and challenges," Journal of Computer Assisted Learning, vol. 26, no. 6, pp. 549–562, 2010.

(BBT nhận bài: 09/05/2016, phản biện xong: 25/05/2016)