NGHIÊN CỨU MỘT SỐ ĐẶC ĐIỂM CÂU HỎI TRẮC NGHIỆM NHIỀU LỰA CHỌN BẰNG LÝ THUYẾT ĐÁP ỨNG CÂU HỎI

Nguyễn Thuỳ Linh, Bùi Mỹ Hạnh
Trường Đại Học Y Hà Nội

Lý thuyết trắc nghiệm hiện đại (lý thuyết đáp ứng câu hỏi của Rasch) hiện nay được áp dụng rộng rãi để đánh giá mối tương quan giữa năng lực của thí sinh với độ khó của từng câu hỏi. Đây là một công cụ mới trong đo lường, đánh giá kết quả học tập bằng hình thức trắc nghiệm khách quan. ***Mục tiêu****: (1) Xác định hình thức, điểm số bài thi trắc nghiệm. (2) Phân tích tương tác giữa câu hỏi với mô hình lý thuyết đáp ứng câu hỏi. 3. Phân tích sự hoà hợp giữa thí sinh với mô hình.* ***Đối tượng và phương pháp nghiên cứu:*** *402 bài thi trắc nghiệm của sinh viên Y Hà Nội năm thứ 2. Sử dụng phần mềm QUEST dựa trên mô hình Rasch để phân tích những tham số đặc trưng câu hỏi – thí sinh và sự hòa hợp của thí sinh với mô hình trong từng đề thi.* ***Kết quả và bàn luận:*** *đa số các câu hỏi MCQ đều phù hợp với mô hình Rasch, câu 22, 25, 30, 31, 44 - đề 1 không phù hợp với mô hình. Độ phân biệt của các câu hỏi dao động từ thấp đến cao, đa số câu hỏi có độ phân biệt cao, cả 2 đề có những câu có độ phân biệt thấp, đặc biệt câu 31, 44 đề 1, câu 42, 47 đề 2 có độ phân biệt âm. Các phương án gây nhiễu tốt tuy nhiên có một số phương án gây nhiễu không có khả năng gây nhiễu. Đa số năng lực thí sinh đều phù hợp với độ khó câu hỏi. Một số thí sinh có đặc điểm bất thường (làm sai câu dễ hơn và/hoặc làm đúng câu khó hơn). Các kết quả này sẽ là cơ sở giúp cho việc xem xét, chỉnh sửa, chuẩn hóa ngân hàng câu hỏi ngày một tốt hơn. Tìm hiểu mối tương quan cụ thể của từng thí sinh - từng câu hỏi sẽ giúp cho việc đánh giá công bằng và chính xác.* ***Kết luận:*** *mô hình Rasch hoàn toàn thích hợp để phân tích tương tác giữa câu hỏi với mô hình và sự hoà hợp của thí sinh với mô hình.*

Từ khóa: câu hỏi trắc nghiệm, rasch

I. ĐẶT VẤN ĐỀ

Trắc nghiệm khách quan (TNKQ) là một công cụ quan trọng giúp đánh giá chính xác, toàn diện, khách quan kết quả học tập và trình độ của sinh viên, học sinh. Ở Việt Nam, việc áp dụng TNKQ vào thực tiễn hoạt động đánh giá còn diễn ra rất chậm và chưa đi vào thường quy. Trong đào tạo chuyên ngành Y, mỗi sinh viên Y cũng phải trải qua nhiều hình thức lượng giá như: quan sát, vấn đáp, tự luận, thi trắc nghiệm (TN)... Loại hình thi bằng các câu hỏi trắc nghiệm cũng được khuyến khích sử dụng nhưng chưa phổ cập, việc công bố ngân hàng câu hỏi cũng hạn chế do chưa đủ cả về số lượng và chất lượng câu hỏi. Làm thế nào có được những ngân hàng câu hỏi (NHCH) đảm bảo tính tin cậy và tính giá trị cho đánh giá kết quả học tập của sinh viên một cách công bằng, khách quan? Đây là một quy trình đòi hỏi tuân thủ khắt khe những nguyên tắc chế tác, chỉnh sửa từng câu, từng ý trong mỗi câu trắc nghiệm dựa trên bằng chứng khoa học về đo lường và đánh giá các câu hỏi trắc nghiệm. Hai lý thuyết trắc nghiệm đang được sử dụng rộng rãi trong nghiên cứu về trắc nghiệm là lý thuyết trắc nghiệm cổ điển và lý thuyết trắc nghiệm hiện đại. Lý thuyết trắc nghiệm hiện đại (lý thuyết đáp ứng câu hỏi của Rasch), cùng với sự hỗ trợ của công nghệ thông tin đã giúp cho việc phân tích, xử lý số liệu trở lên dễ dàng [3]. Ưu điểm của lý thuyết đáp ứng câu hỏi là không phụ thuộc vào mẫu, chỉ phụ thuộc vào khả năng đo lường đã thực sự hứa hẹn là một công cụ mới cho việc xây dựng và phát triển NHCH, phân tích các kết quả TNKQ để hoàn chỉnh NHCH, chủ động thiết kế các đề TN theo mục tiêu mong muốn. Các nhà khoa học cũng đã có nhìn nhận, đánh giá ban đầu về những lợi ích của mô hình này [4]. Cho đến nay, Việt Nam đã có một số nghiên cứu về mô hình

này [2, 5]. Với mong muốn ứng dụng mô hình Rasch để đánh giá bộ câu hỏi trắc nghiệm đối với sinh viên Y2 tại Đại học Y Hà Nội, chúng tôi tiến hành đề tài nhằm **mục tiêu:**

1. Xác định hình thức, điểm số bài thi trắc nghiệm.

2. Phân tích tương tác giữa câu hỏi với mô hình lý thuyết đáp ứng câu hỏi.

3. Phân tích sự hoà hợp giữa thí sinh với mô hình.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

1. Đối tượng nghiên cứu

402 bài thi trắc nghiệm kết thúc học phần lý thuyết của sinh viên Y năm thứ 2 được làm theo 2 đề, 48 câu đề chẵn và 48 câu đề lẻ.

2. Phương pháp nghiên cứu

2.1. Thu thập số liệu: kết quả bài làm của 402 sinh viên Y2 được nhập bằng chương trình SPSS13.0.

2.2. Xử lý số liệu: kết quả bài làm thí sinh sau khi được nhập liệu bằng chương trình SPSS 13.0 sẽ được chuyển sang phần mềm QUEST của ACER (Úc) để phân tích theo mô hình Rasch. Mô hình Rasch mô tả mối tương tác giữa từng TS với từng câu hỏi của đề TN. Mô hình này dựa trên giả thiết "một TS có khả năng cao hơn một TS khác phải có một xác suất lớn hơn để trả lời đúng một câu bất kỳ, tương tự, một câu hỏi khó hơn một câu hỏi khác thì đối với bất kỳ TS nào xác suất trả lời đúng câu sau là cao hơn câu trước". Mục đích chung của CHTN là đo lường một khả năng nào đó và có thể mô tả bằng toán học mối liên hệ giữa khả năng của một người và thành tích đáp ứng của người đó. Với kỳ thi TN có 96 CH MCQ và 402 SV tham gia, mỗi câu hỏi có một độ khó và mỗi TS có một năng lực nhất định thì sẽ có 402 x 96 = 38592 mối tương tác giữa TS và câu hỏi. Nhập các đáp án, phần mềm Quest sẽ cho ra các số liệu về: điểm bài TN theo đáp án đã cho; Độ tin cậy của bài TN; Khả năng của thí sinh; Phân tích các câu TN: độ khó (Percent), độ phân biệt (Pt - biserial), độ tin cậy thống kê của độ phân biệt (P - value).v.v. của các câu trả lời đúng và câu mồi; phân tích sự hoà hợp của các câu TN và khả năng thí sinh.

3. Các thông số nghiên cứu

3.1. Hình thức chung – Điểm thô của bài trắc nghiệm

3.2. Phân tích tương tác giữa câu hỏi với mô hình Rasch

- Sự hòa hợp của câu hỏi với mô hình.

- Sự phân bố độ khó câu hỏi và khả năng của thí sinh.

- Độ khó, độ phân biệt, tính tin cậy thống kê của từng câu hỏi.

3.3. Phân tích sự hòa hợp thí sinh – mô hình

- Sự hòa hợp thí sinh - mô hình ở các đề thi.

- Những đặc điểm bất thường về tương tác thí sinh - câu hỏi.

III. KẾT QUẢ

1. Hình thức và điểm thô trung bình

1.1. Hình thức chung

Đề thi gồm hai đề chẵn - lẻ, mỗi đề có 48 câu hỏi MCQ, mỗi câu có 5 phương án lựa chọn A, B, C, D, E. Nếu chọn phương án đúng thí sinh được 1 điểm. Điểm toàn bài sẽ là % số câu làm đúng quy về thang điểm 10.

1.2. Điểm thô trung bình của bài trắc nghiệm

Đề 1: n = 200; $\overline{X}$ =56,51 ± 7,55

Đề 2: n = 202; $\overline{X}$ = 54,80 ± 7,09

Nhận xét: điểm thô trung bình của nhóm thí sinh làm đề số 1 cao hơn đề 2 có ý nghĩa thống kê với p < 0,05.

2. Phân tích tương tác giữa câu hỏi với mô hình Rasch

2.1. Sự hòa hợp của câu hỏi với mô hình

Bảng.1a. Sự hòa hợp của câu hỏi đề 1 với mô hình

INFIT MNSQ	.63	.71	.83	1.00	1.20	1.40	1.60
1 item 1			.	|	*	.	
2 item 2			.	| *		.	
3 item 3			.	|	*	.	
4 item 4			.	| *		.	
5 item 5			.	*		.	
22 item 22		*	.	|		.	
25 item 25		*	.	|		.	
30 item 30		*	.	|		.	
31 item 31			.	|		. *	
44 item 44			.	|		.	*

Bảng 1b. Sự hòa hợp của câu hỏi đề 2 với mô hình

INFIT MNSQ	.63	.71	.83	1.00	1.20	1.40	1.60
1 item 1			.	*		.	
2 item 2			.	| *		.	
3 item 3			.	|	*	.	
4 item 4			.	|	*	.	
5 item 5			.	| *		.	
6 item 6			.	*		.	
7 item 7			.	* |		.	
8 item 8			.	* |		.	
9 item 9			.	* |		.	
10 item 10			.	* |		.	

Trong bảng 1a, câu 22, 25, 30, 31 và 44 của đề 1 không hòa hợp với mô hình.

2.2. Sự phân bố độ khó câu hỏi và khả năng của thí sinh (bảng 2)

2.3. Phân tích độ khó, độ phân biệt, các mồi nhử câu hỏi trắc nghiệm

Độ phân biệt của các câu hỏi dao động từ thấp đến cao, đa số câu hỏi có độ phân biệt cao, cả 2 đề có những câu có độ phân biệt thấp (12 câu đề 1, 12 câu đề 2), đặc biệt câu 31, 44 đề 1, câu 42, 47 đề 2 có độ phân biệt âm.

Phân tích một vài ví dụ minh họa cụ thể từng câu hỏi ta có kết quả sau:

- Ví dụ 1: câu có độ phân biệt cao > 0,3.

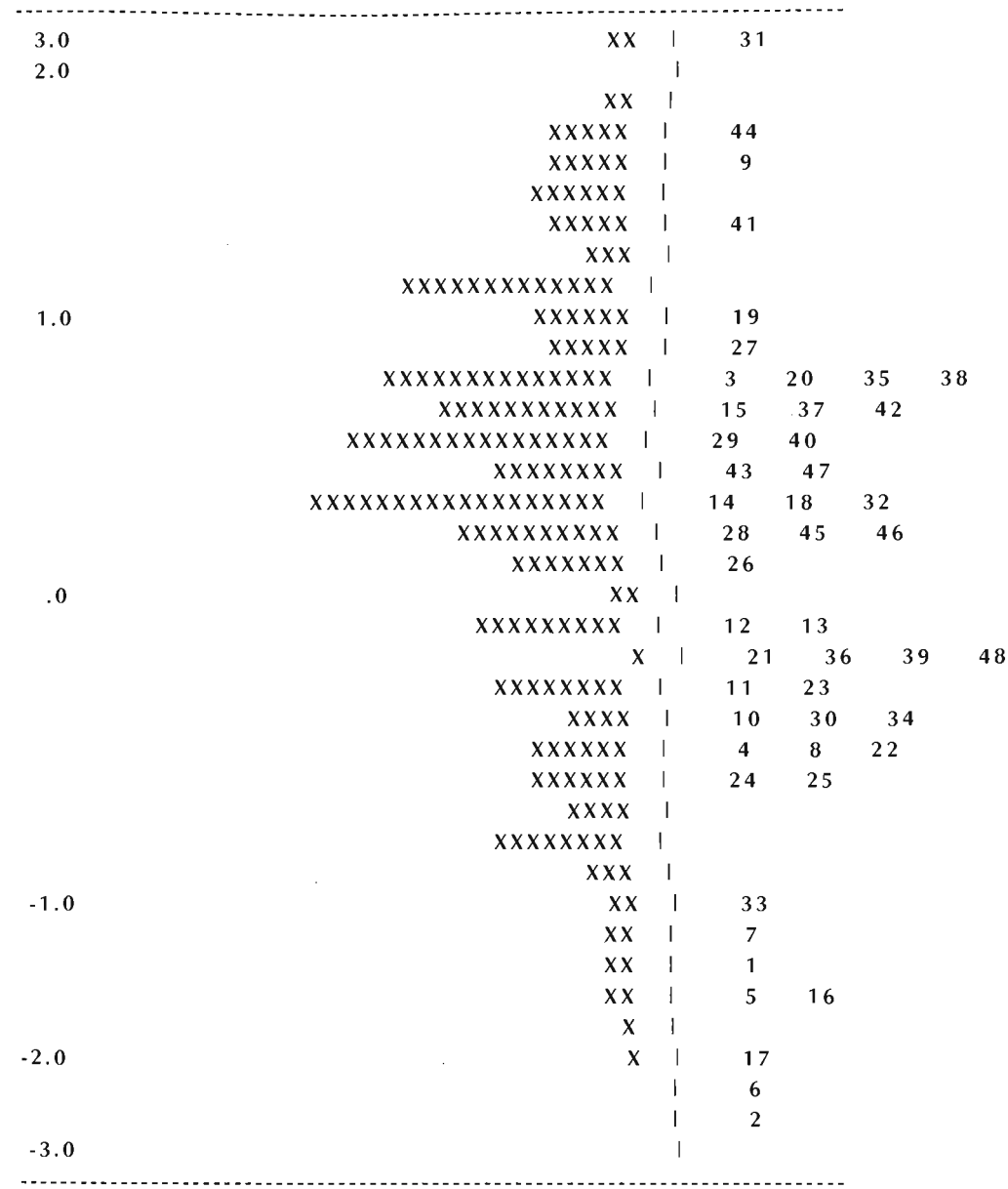
Nhận xét 1: đây là một câu có độ phân biệt cao (0,55) và độ khó vừa phải (50%).

- Ví dụ 2: câu có độ phân biệt trung bình từ 0,2 - 0,29.

Nhận xét 2: đây là câu có cả độ phân biệt (0,22) và độ khó trung bình (54,8%).

- Ví dụ 3: câu có độ phân biệt thấp < 0,2

Bảng 2. Phân bố độ khó câu hỏi và khả năng của thí sinh đề 1



Ví dụ 1:

Câu 18	Đề 1		Độ phân biệt = .55			
Phương án	A	B	C*	D	E	Số SV không làm
Số SV chọn	9	6	99	40	44	2
Độ khó	4.5	3.0	50.0	20.2	22.2	
Độ phân biệt	-.18	-.02	.55	.04	-.61	
Giá trị p	.005	.369	.000	.276	.000	

Ví dụ 2:

Câu 26 Đề 1 — Độ phân biệt = .22

Phương án	A	B	C	D*	E	Số SV không làm
Số SV chọn	55	14	17	109	4	1
Độ khó	27.6	7.0	8.5	54.8	2.0	
Độ phân biệt	.04	-.14	-.31	.22	-.04	
Giá trị p	.312	.024	.000	.001	.280	

Ví dụ 3:

Câu 2 Đề1 — Độ phân biệt = .06

Phương án	A*	B	C	D	E	missing
Số SV chọn	191	9	0	0	0	0
Độ khó	95.5	4.5	.0	.0	.0	
Độ phân biệt	.06	-.06	NA	NA	NA	
Giá trị p	.185	.185	NA	NA	NA	

Câu 9 Đề — Độ phân biệt = .05

Phương án	A*	B	C	D	E	missing
Số SV chọn	51	149	0	0	0	0
Độ khó	25.5	74.5	.0	.0	.0	
Độ phân biệt	.05	-.05	NA	NA	NA	
Giá trị p	.256	.256	NA	NA	NA	

Nhận xét 3: câu có độ phân biệt rất thấp mà có thể là rất dễ (câu 2 đề 1), hoặc rất khó (câu 9 đề 1).

- Ví dụ 4: Câu có độ phân biệt âm (< 0).

Câu 44 Đề 1 — Độ phân biệt = -.39

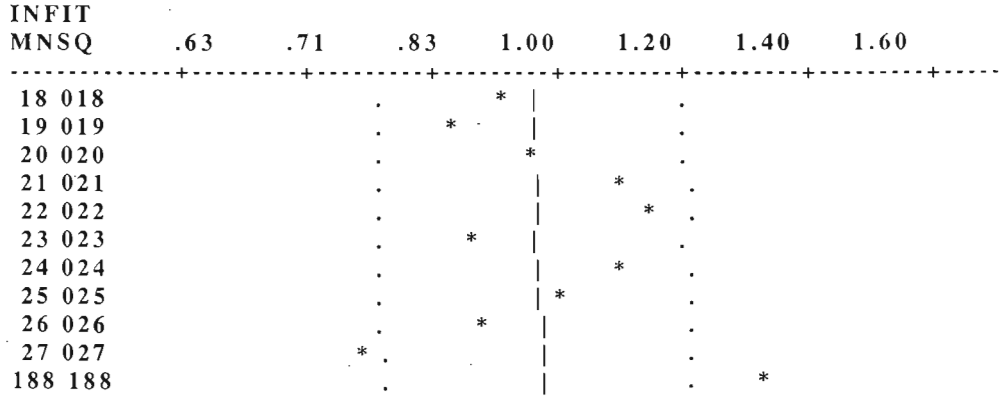
Phương án	A	B*	C	D	E	Số SV không làm
Số SV chọn	23	46	32	83	4	12
Độ khó	12.2	24.5	17.0	44.1	2.1	
Độ phân biệt	.19	-.39	.14	.09	.05	
Giá trị p	.004	.000	.028	.101	.245	

Nhận xét 4: câu 31, 44 đề 1 là câu có độ phân biệt âm, thí sinh có điểm toàn bài cao chọn đáp án đúng ít hơn những thí sinh có điểm toàn bài kém.

3. Phân tích sự hòa hợp thí sinh – mô hình

3.1. Sự hòa hợp thí sinh - mô hình ở các đề thi

Bảng 3. Sự hòa hợp thí sinh – mô hình ở đề 1



Nhận xét: một số sinh viên có đáp ứng bất thường như thí sinh thứ 27, 188.

3.2. Những đặc điểm bất thường về tương tác thí sinh - câu hỏi

Với kỳ thi ở môn học này, chúng ta sẽ thu được 402 kết quả đánh giá sự phù hợp khả năng thí sinh và độ khó câu hỏi như một ví dụ minh họa sau:

Bảng 4. Phân bố tương quan giữa độ khó và kết quả bài làm của thí sinh 188 - Đề 1

```
------------------------------ K I D M A P------------------------------
---Khó hơn làm đúng-- ---------- Khó hơn làm sai - ----------
                        |  |
                        |  |  31(E)
                   44   |  |
                        |  |  9(B)
                        |  |  41(E)
                   27   |  |  19(C)
                    3   |  |  20(C)  38(D)
               37   15  |  |  35(E)  42(E)
                        |  |  29(C)  40(C)  47(A)
                        |  |  14(B)  18(E)  32(A)  43(E)
                   28   |  |  45(D)  46(C)
                        |  |  26(B)
                        |  |  12(B)  13(B)
                        |  |  21(A)  36(B)  39(B)  48(A)
                        |  |  11(B)  23(E)
               10    8  |  |  22(C)  30(D)  34(E)
                        |  |  4(B)  25(C)
.....                        33(E)
                        |  |  7(B)
                    1   |  |
                        |  |  5(B)  16(D)
                   17    .....
                        |  |  6(B)
                        |  |  2(B)
                        |  |
-----Dễ hơn làm đúng ------- ---------- Dễ hơn làm sai -- ----------
==========================================================
```

Với thí sinh SBD 188, chỉ có 2 câu đúng trong mức dễ hơn - làm đúng, câu 2 là câu rất dễ nhưng lại làm sai trong khi câu 44 là câu rất khó lại làm đúng.

IV. BÀN LUẬN

1. Điểm thô trung bình của bài trắc nghiệm

Điểm thô trung bình của nhóm thí sinh làm đề 1 cao hơn so với điểm thô trung bình của đề 2 có ý nghĩa thống kê ($p < 0,05$). Theo quan điểm thống kê thông thường, đây là hai đề có độ khó khác nhau và/hoặc nhóm thí sinh làm đề 1 có khả năng hơn nhóm thí sinh làm đề 2. Tuy nhiên, trong lĩnh vực đo lường giáo dục, cách tính toán điểm thô trung bình rồi đem so sánh chưa nói lên được gì nhiều. Điểm thô này sẽ không thể kết luận sinh viên đó giỏi hay kém về môn học nào vì thiếu chuẩn chung về độ khó hay khả năng của một nhóm chuẩn. Ngay trong một môn học, nếu ra những đề thi có độ khó các câu hỏi khác nhau thì điểm thô lại càng ít có giá trị trong đánh giá thành quả của thí sinh. Mô hình Rasch sẽ giúp ta có những thông số để giải quyết những vấn đề này [10].

2. Phân tích tương tác giữa câu hỏi với mô hình Rasch

2.1. Sự hòa hợp của câu hỏi với mô hình

Sự hoà hợp của các câu hỏi ở 2 đề được thể hiện chi tiết trong kết quả phân tích từ phần mềm QUEST. 48 câu hỏi ở đề 2 đều nằm phía trong của 2 đường chấm là phù hợp với mô hình. Ở đề 1, có một số câu hỏi không phù hợp với mô hình là câu 22, 25, 30, 31 và câu 44. Đây là những câu hỏi không nên dùng tiếp trong các kỳ thi sau hoặc cần chỉnh sửa trước khi sử dụng lại.

2.2. Sự phân bố độ khó câu hỏi và khả năng của thí sinh

Kết quả trình bày ở bảng 2 cho biết sự phân bố độ khó câu hỏi tương ứng với sự phân bố khả năng của mỗi nhóm thí sinh làm đề 1 trên cùng một thang đo. Đường vạch rời chính giữa là thang đo, dấu x là phân bố các thí sinh theo mức năng lực, bên phải là phân bố các câu hỏi theo độ khó. Như vậy, trên cùng một thang đo có thể thấy khái quát mức năng lực chung của nhóm thí sinh cao hơn với độ khó trung bình của đề thi cũng như những câu hỏi quá dễ (câu hỏi 2, 6, 17) và câu quá khó như câu 31, 44.

2.3. Phân tích độ khó, độ phân biệt, các mồi nhử câu hỏi trắc nghiệm

Độ phân biệt của câu hỏi cho biết liệu có đúng những học sinh có điểm toàn bài cao có xu hướng làm đúng câu hỏi khó hơn những học sinh có điểm toàn bài kém. Theo Dương Thiệu Tống và một số tác giả tổng kết, độ phân biệt ≥ 0,3 được coi là tốt, từ 0,2 - 0,29 được coi là tạm được và dưới 0,2 (đặc biệt là < 0) thì cần xem xét và loại bỏ. Ví dụ câu 18 đề 1 là một câu có độ phân biệt cao (0,55) và độ khó vừa phải (50%), các phương án gây nhiễu đều có khả năng hấp dẫn thí sinh lựa chọn. Đây là một câu hỏi tốt và nên tái sử dụng trong những kỳ thi tiếp theo.

Thông thường các phương án nhiễu phải có giá trị phân cách là âm nhưng trong câu 26 đề 1, phương án nhiễu A ở lại có giá trị dương. Nếu đây là một kiến thức sai nhưng xu hướng được những sinh viên có điểm toàn bài cao cho là đúng thì việc rà soát lại câu hỏi, tài liệu giảng, ý giảng trên lớp từ phía các bộ môn vẫn rất cần thiết trong trường hợp này.

Một số câu có độ phân biệt rất thấp, có thể là câu rất dễ (câu 2 đề 1), hoặc rất khó (câu 9 đề 1). Ở câu 9 đề 1, sự lựa chọn phương án sai là B chiếm 74,5% cao hơn phương án đúng là A. Các phương án nhiễu C, D, E không có thí sinh nào lựa chọn cần loại bỏ vì không có khả năng đánh lừa được sinh viên [1, 3, 4]. Độ phân biệt của một câu hỏi hoặc một đề trắc nghiệm rõ ràng có liên quan đến độ khó.

Câu 31, 44 đề 1, câu 42, 47 đề 2 là câu có độ phân biệt âm có nghĩa là thí sinh có điểm toàn bài cao chọn đáp án đúng ít hơn những thí sinh có điểm toàn bài kém. Xem xét về độ khó thì đây lại là những câu nằm ở vị trí có độ khó cao, đáp án đúng có giá trị âm trong khi đó các phương án gây nhiễu lại có giá trị dương, do vậy việc không có sự hoà hợp với mô hình cũng là hợp lý và cần loại bỏ.

3. Phân tích sự hòa hợp thí sinh – mô hình

3.1. Đa số năng lực các thí sinh đều phù hợp với mô hình

Đa số năng lực các thí sinh đều phù hợp với mô hình (nằm trong sự hòa hợp), phần mềm QUEST nhanh chóng cho ta xác định được những sinh viên có đáp ứng bất thường như TS thứ 27, TS thứ 188 (đề 1). Đây là một trong những tính năng vượt trội của mô hình Rasch trong đánh giá khả năng của thí sinh. Như vậy, sau khi phần mềm phân tích và xử lý số liệu, chúng ta sẽ thu được 402 kết quả đánh giá sự phù hợp khả năng của TS và độ khó của câu hỏi.

3.2. Những đặc điểm bất thường về tương tác thí sinh - câu hỏi

Từ kết quả trên tất cả những thí sinh có dấu hiệu bất thường, được coi là không hòa hợp sẽ được xem xét cụ thể qua kết quả phân tích sự hòa hợp từng thí sinh với mô hình. Theo lý thuyết, đa số thí sinh đều có đáp ứng bình thường là với câu hỏi Khó hơn sẽ Làm sai và Dễ hơn là Làm đúng; một số thí sinh được coi là có kết quả bất thường nếu câu hỏi Khó hơn – Làm đúng và Dễ hơn – Làm sai. Đây là một trong những tính năng vượt trội của mô hình Rasch trong đánh giá khả năng của mỗi thí sinh trong từng kỳ thi.

Với thí sinh SBD 188, chỉ có 2 câu đúng trong mức dễ hơn - làm đúng, câu 2 là câu rất dễ nhưng

lại làm sai trong khi câu 44 là câu rất khó lại làm đúng. Trường hợp thí sinh không làm được câu dễ hơn có thể xảy ra hai khả năng: Một là thí sinh học lệch, học tủ, hai là tài liệu học tập có thể có những điểm chưa rõ. Trường hợp thí sinh có khả năng thấp hơn nhưng lại làm được những câu hỏi khó có thể do thí sinh quay cóp hoặc đoán mò, cũng có thể thí sinh đó có khả năng "đặc biệt" về lĩnh vực đó. Tình huống thứ ba gợi cho chúng ta ý tưởng phải chăng có thể ứng dụng mô hình này trong việc định hướng cho sinh viên theo những chuyên ngành họ có khả năng tốt hơn những khả năng thông thường [9]? Chính vì điều này, nhiều tác giả đã cho rằng mô hình Rasch là một trong những hướng nghiên cứu đầy hứa hẹn về cung cấp thông tin cho đánh giá hướng nghiệp theo khả năng của con người. Dựa vào kết quả này, giảng viên có thể xem xét bài làm của các sinh viên cá biệt để tìm hiểu nguyên nhân của các câu trả lời bất thường trên [5].

V. KẾT LUẬN

Lý thuyết đáp ứng câu hỏi của Rasch hoàn toàn thích hợp để phân tích tương tác giữa câu hỏi với mô hình.

Sử dụng mô hình Rasch có thể đánh giá sự hòa hợp của thí sinh với mô hình và phát hiện những đáp ứng bất thường trong mối tương tác giữa khả năng thí sinh và độ khó câu hỏi.

KHUYẾN NGHỊ

Ứng dụng mô hình Rasch để hoàn chỉnh bộ câu hỏi trắc nghiệm phù hợp với mục đích của từng kì thi.

TÀI LIỆU THAM KHẢO

1. Nguyễn Phụng Hoàng và **Võ Ngọc Lan (1996)** Phương pháp trắc nghiệm trong kiểm tra và đánh giá thành quả học tập, Nhà xuất bản giáo dục, tr. 219.

2. Nguyễn Hoàng Minh, Phạm Nhật Minh, Nguyễn Thị Luyến, Lưu Ngọc Hoạt, Bùi Mỹ Hạnh (2007) Ứng dụng phần mền Quest theo mô hình Rasch trong phân tích bộ câu hỏi thi trắc nghiệm môn Sinh lý học, Tạp chí Y học thực hành, 606 – 607: tr. 614 – 621.

3. Dương Thiệu Tống (2005) Trắc nghiệm và đo lường thành quả học tập, nhà xuất bản Khoa học - Kỹ thuật.

4. Lâm Quang Thiệp (2006) Việc áp dụng những thành tựu hiện đại về đo lường giáo dục ở nước ta, Tạp chí Giáo dục số133, tr12 - 13

5. Nguyễn Bảo Hoàng Thanh (2008) Sử dụng phần mềm Quest để phân tích câu hỏi trắc nghiệm khách quan, Tạp chí khoa học công nghệ, Đại học Đà Nẵng số 2.

6. Downing SM. (2003) Item response theory: applications of modern test theory in medical education, Med Educ. 37(8): pp. 739 - 45.

7. Hays RD, Liu H, Spritzer K, Cella D. (2007) Item response theory analyses of physical functioning items in the medical outcomes study, Med Care., 45 (5 Suppl 1): S32 - 8.

8. Ivailo Partchev (2004) A visual guide to item response theory, Friedrich - Schiller - University Jena, 61p.

9. Lawson DM. (2006) Applying the item response theory to classroom examinations. J Manipulative Physiol Ther., 29(5): pp. 393 - 7.

10. Raymond J. Adams, Siek - Toon Khoo (1993) - QUEST - The Interactive Test Analysis System - ACER, Australia.

Summary

STUDY ON CHARACTERISTICS OF MULTIPLE CHOICE QUESTIONS BY RESPONSE QUESTION THEORY

Modern test theory is now widely applied to assess the correlation between the ability of candidates and the difficulty of each question. **Objectives:** (1) Defining the form and marks of the test. (2) Analyzing the correlation between the questions and the candidates based on response question theory model. 3. Analyzing the harmony between the candidates and the model. **Method:** 402 tests of Hanoi medical 2th year students. **Result:** Most MCQs fit Rasch model. Most questions have high distinction level. Most abilities of candidates are suitable for question difficulty. Some candidates have unusual features. Measurement capabilities of candidates with difficulty level, distinction level of each question will help to review, edit, standardize question banks better and better. **Conclusion:** Rasch model definitely fits to analyze the distribution of question difficulty, evaluating the specific parameters of questions, topic test. Model also detects the unusual response in correlation of candidate ability and question difficulty.

Keywords: response question theory, standardize question banks

ỨNG DỤNG MÔ HÌNH HAI THAM SỐ TRONG NGHIÊN CỨU MỘT SỐ ĐẶC TÍNH TRẮC NGHIỆM TÂM LÝ ĐÁNH GIÁ TRẠNG THÁI TÂM TRÍ TỐI THIỂU (MMSE) Ở NGƯỜI BÌNH THƯỜNG TUỔI 50 - 59

Trần Thúy Liễu[1], Bùi Mỹ Hạnh[2]

[1]Trường Cao đẳng Y tế Hà Nội; [2]Trường Đại học Y Hà Nội

Trắc nghiệm đánh giá trạng thái tâm trí tối thiểu (Mini Mental State Examination - MMSE) được chuyển thể sang nhiều phiên bản với những ngôn ngữ khác nhau trong đó có Việt Nam và được sử dụng như một công cụ sàng lọc chức năng nhận thức. ***Mục tiêu:*** *ứng dụng mô hình hai tham số để tìm ra một số đặc tính các câu hỏi và toàn bộ bộ trắc nghiệm MMSE trên người bình thường tuổi 50 - 59.* ***Đối tượng và phương pháp nghiên cứu:*** *120 người tuổi từ 50 - 59, gồm 60 nam và 60 nữ được thực hiện trắc nghiệm MMSE. Các điểm số thu được được đưa vào chương trình BILOG để xử lý và phân tích.* ***Kết quả:*** *sau phase 1: 9/30 câu bị loại bỏ khỏi chương trình xử lý do không phù hợp (hai câu có độ phân biệt < 0, bảy câu có độ phân biệt bằng 0). Sau phase 2: trong hai mươi mốt câu được phân tích mười câu có độ khó không phù hợp, mười một câu có độ khó phù hợp. Điểm thô tối thiểu của 21 câu được xác định từ hàm thông tin câu hỏi là 18.* ***Bàn luận:*** *trắc nghiệm MMSE có nhiều câu tỏ ra quá dễ so với năng lực của đối tượng, chỉ có một số ít câu cần thiết thể hiện năng lực nhận thức của đối tượng nghiên cứu.* ***Kết luận*** *sử dụng mô hình hai tham số cho thấy có thể đơn giản hóa bộ trắc nghiệm MMSE, việc loại bỏ một số câu không phù hợp không những làm công cụ này ngắn lại, dễ tiến hành, mà còn duy trì được tính giá trị và độ tin cậy của trắc nghiệm để có thể sử dụng như một test sàng lọc ở cộng đồng.*

Từ khóa: trắc nghiệm tâm lý, trạng thái tâm trí tối thiểu