

SỬ DỤNG PHẦN MỀM QUEST ĐỂ PHÂN TÍCH CÂU HỎI TRẮC NGHIỆM KHÁCH QUAN

USING QUEST SOFTWARE TO ANALYZE OBJECTIVE TEST QUESTIONS

NGUYỄN BẢO HOÀNG THANH
Trường Đại học Sư phạm, ĐHQĐ

TÓM TẮT

Bài báo này trình bày các kết quả khi sử dụng phần mềm QUEST để phân tích một bài kiểm tra bằng phương pháp trắc nghiệm. Quest là một chương trình phân tích và đánh giá câu hỏi trắc nghiệm, bài trắc nghiệm, được xây dựng dựa trên lý thuyết đáp ứng câu hỏi IRT

ABSTRACT

The report presents the results of using Quest software to analyze a check-up through objective test. Quest is a program which analyzes and evaluates multiple choice questions, multiple choice task created based on Item response theory.

1/ Trong những năm gần đây, Bộ Giáo dục và Đào tạo đã, đang thực hiện các chủ trương đổi mới nội dung chương trình giảng dạy, cải tiến phương pháp thi, kiểm tra đánh giá kết quả học tập của học sinh, sinh viên từ phổ thông đến đại học. Phương pháp trắc nghiệm khách quan được đưa vào trong các kỳ kiểm tra, thi học kỳ, tuyển sinh, tốt nghiệp v.v... nên sự quan tâm của các giảng viên và các cơ quan quản lý giáo dục ngày càng nhiều. Do đó việc nghiên cứu sử dụng phần mềm Quest để phân tích CHTN là cần thiết.

2/ Để viết câu hỏi trắc nghiệm (CHTN) và xây dựng ngân hàng CHTN dùng để kiểm tra đánh giá (KTĐG), thi học kỳ đạt yêu cầu của mục tiêu dạy học thì phải nắm vững qui trình viết và phân tích thống kê từng câu hỏi TN, cũng như trên bài TN. Nguyên tắc chung để phân tích CHTN, một bài TN là chúng ta thường so sánh câu trả lời của sinh viên (SV) ở mỗi câu hỏi với điểm số chung của toàn bài với mong muốn có nhiều SV khá giỏi trả lời đúng, đồng thời có ít SV yếu trả lời được câu hỏi đó, nghĩa là phổ các điểm của một lớp SV trải càng rộng càng tốt. Việc phân tích thống kê CHTN nhằm xác định các chỉ số như: độ khó, độ phân biệt, hệ số độ tin cậy... của CHTN. Theo lý thuyết trắc nghiệm cổ điển thì sau khi chấm điểm bài TN xong, ta thường chia lớp làm 3 nhóm: Nhóm điểm cao chọn 27% SV đạt điểm cao nhất và nhóm điểm thấp chọn 27% SV điểm thấp nhất. Từ đó tính độ khó, độ phân biệt của CHTN, của bài trắc nghiệm. Phương pháp ở trên đơn giản mà mỗi giảng viên ở lớp đều có thể tính với từng câu, từng bài TN. Nhưng nhược điểm của phương pháp này là chỉ chọn 54% SV trong hai nhóm điểm cao và điểm thấp, còn lại 46% nhóm điểm ở mức trung bình không được xét đến. Từ sau thập niên 70 đến nay một lý thuyết phân tích câu hỏi trắc nghiệm hiện đại ra đời dựa trên thuyết đáp ứng câu hỏi của Georg Rasch thường gọi là lý thuyết IRT (Item response theory). Thuyết IRT được phát triển rất nhanh nhờ khả năng tính toán bằng máy vi tính.

Để khắc phục nhược điểm nêu trên, Hội đồng nghiên cứu giáo dục Australia dựa trên thuyết IRT đã xây dựng phần mềm Quest để phân tích số liệu thống kê của câu TN và bài TN. Thuyết đáp ứng câu hỏi của Rasch mô hình hoá mối liên hệ giữa mức độ khả năng

của người làm TN và đáp ứng của người ấy với một câu TN. Mỗi câu TN được mô tả bằng một thông số (độ khó) ký hiệu là δ và mỗi người làm TN được mô tả cũng bằng một thông số (khả năng) ký hiệu là β . Mỗi khi một người cố gắng trả lời một câu hỏi, các thông số độ khó và khả năng tác động lẫn nhau, để cho một xác suất đáp ứng của người làm trắc nghiệm ấy. Dạng toán học của mô hình này là:

$$P_{nil} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)} \quad (7,205)$$

Trong đó P_{nil} là xác suất của một người với khả năng β_n làm đúng câu trắc nghiệm có độ khó δ .

3/ Phân tích các chỉ số thống kê theo QUEST.

Sau khi thi xong, các phiếu thi được đưa vào máy quét OFSCAN hoặc nhập từng bài trắc nghiệm vào máy tính. Chương trình Quest sẽ cho ta các bảng số liệu sau đây: Chấm điểm bài TN theo đáp án đã cho - Độ tin cậy của bài trắc nghiệm - Khả năng của người làm trắc nghiệm - Phân tích các câu TN, tính độ khó, độ phân biệt Rpbis, P-Value: độ tin cậy thống kê của độ phân biệt v.v. của các câu lựa chọn, cả câu lựa chọn đúng lẫn các câu môi (bảng 1) - Phân tích sự hoà hợp của các câu TN và khả năng người làm TN, khả năng của mỗi người (bảng 2, bảng 3) cùng với các trường hợp bất thường của người ấy, nếu có, bằng cách nêu ra các câu TN vượt quá khả năng mà người ấy làm đúng hoặc các câu dưới khả năng mà người ấy làm sai.v.v. có thể do người này quay cốp, đoán mò.v.v. từ đó xem lại quá trình học tập của người này (bảng 4, bảng 5). [5,16-40), [6,124-140].

4/ Trong quá trình xây dựng ngân hàng câu hỏi trắc nghiệm môn vật lý đại cương 1 dùng cho SV khối Đại học kỹ thuật, chúng tôi đã dùng phần mềm Quest để phân tích các CHTN và các bài thi để lựa chọn các câu hỏi đạt yêu cầu lưu vào ngân hàng CHTN.

Sau đây là các bảng mô tả kết quả phân tích bài kiểm tra vật lý 1 gồm 40 câu do 108 sinh viên khoá 05 trường Đại học Bách khoa - Đại học Đà Nẵng thực hiện.

Bảng 1: Các chỉ số thống kê các câu hỏi trắc nghiệm đề 16.

Run ONE:DE SO :16 TNSP DOT 2 10LOP K 05 . Test

```
-----
Item Analysis Results for Observed Responses              15/
2/2008 8:42
all on all (N = 108 L = 40)
-----
Item      1: item 1                                     (key=B)
Categories      A          B*          C          D          E
missing
Count           11          75          7          3          9
3
Percent (%)     10.5         71.4         6.7         2.9         8.6
Pt-Biserial    -.22          .25          .03          .00         -.20
p-value        .013         .004         .381         .487         .021
```

Item 39: item 39 (key=A)

Categories	A*	B	C	D	E
missing					
Count	62	13	11	9	9
4					
Percent (%)	59.6	12.5	10.6	8.7	8.7
Pt-Biserial	.46	-.29	-.15	-.07	-.23
p-value	.000	.002	.062	.227	.010

Item 40: item 40	(key=B)				
Categories	A	B*	C	D	E
missing					
Count	20	42	12	14	12
8					
Percent (%)	20.0	42.0	12.0	14.0	12.0
Pt-Biserial	-.05	.31	-.07	-.28	-.04
p-value	.296	.001	.249	.003	.363

Trong đó: Item: Câu hỏi số; Categories: Câu chọn; câu nào có dấu (*) là câu đúng, còn lại là câu nhiễu; Count: Số SV chọn câu tương ứng. Percent (%): Phần trăm số SV chọn câu trả lời tương ứng; P_t-biserial = R_{pbis}: Độ phân biệt câu TN, P-Value: Độ tin cậy thống kê của độ phân biệt; Missing: Số SV không tìm được câu chọn, bỏ trống câu đó.

Dựa vào bảng phân tích này ta có cơ sở khoa học để lựa chọn các CHTN có độ khó, độ phân biệt, độ tin cậy đạt yêu cầu để lưu vào ngân hàng CHTN dùng cho các lần sau, ví dụ như câu 40, có độ khó là 0,42, độ phân biệt là 0,31 các câu nhiễu A, C, D, E có R_{pbis} âm, cho thấy các câu nhiễu tốt, có nhiều SV nhóm yếu chọn hơn SV nhóm giỏi. Đây là CHTN đạt yêu cầu.

Bài trắc nghiệm này có điểm trung bình là 22,95 - Độ tin cậy 0,82

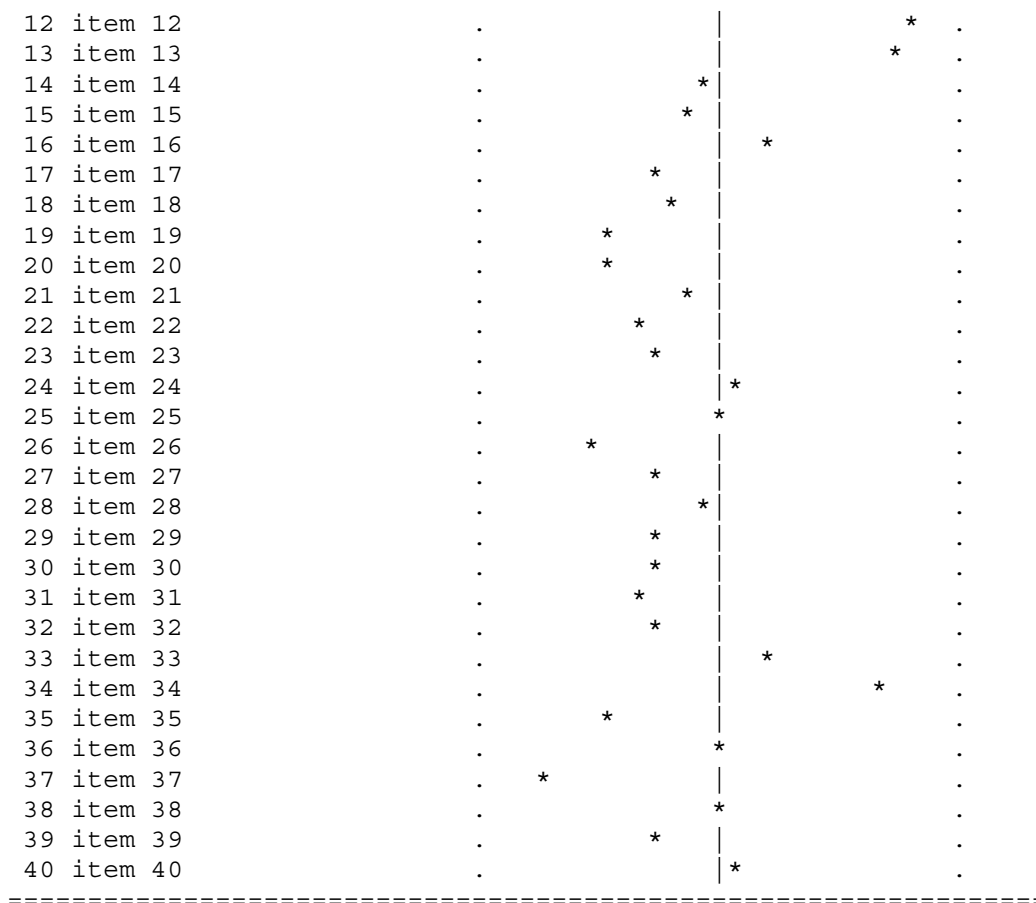
Bảng số 2: Minh họa sự phù hợp các câu TN trong bài TN đề số 16

Run ONE:DE SO :16 TNSP DOT 2 10LOP K 05 . Test

```

-----
Item Fit                                     15/
2/2008 8:42
all on viet (N = 108 L = 40)
-----
-----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        .          |          *          .
11 item 11        .          |          *          .
-----

```



Trong biểu đồ Item fit, mỗi câu TN biểu thị bằng (*) có 40 CHTN đều nằm ở phía trong của 2 đường chấm thẳng đứng có giá trị trung bình bình phương độ hoà hợp (infit mean square) viết tắt: infit MNSQ là (1-0,3) và (1+0,3) là phù hợp với mô hình Rasch. Nếu có CHTN nào nằm ngoài 2 đường chấm trên, không phù hợp với mô hình thì phải đọc và sửa lại. Biểu đồ này chứng tỏ 40 CHTN này đều phù hợp.

Bảng số 3: Biểu đồ minh hoạ sự phân bố độ khó CHTN với năng lực của sinh viên

Run ONE:DE SO :16 TNSP DOT 2 10LOP K 05 . Test

Item Estimates (Thresholds) 15/ 2/2008 8:42
all on viet (N = 108 L = 40)

4.0

XX

3.0

	XX				
	XX				
2.0	XXX				
	XX				
	X				
	XXX	31			
	XXX	34			
	XXXX				
	XXX	21			
	XXXXXX	4	12	32	37
1.0	XXX				
	XXXXXXXX	28	40		
	XXXXXX	5			
	XXXXXX	15	26		
	XXXXXXXX	16	30	33	36
	XX	6	9	29	38
	XXXXXXXX	18			
.0	XXXXXX	24	39		
	XXXX	7	10	27	
	XXXXX				
	XX	25			
	XXXXX	1	2	13	
	XXX	35			
	XXXX	3	11		
	XX				
-1.0	XX	20	22		
	XX	19			
	X				
		23			
		8			
-2.0					
		14			
		17			
-3.0					

Each X represents 1 students

Biểu đồ Item Estimates trình bày sự phân bố thống kê tương quan giữa năng lực của SV (đấu X) với sự phân bố độ khó của các CHTN. Trong biểu đồ này các số bên phải đường chấm thẳng đứng trình bày sự phân bố các CHTN theo độ khó của từng câu trong bài TN. Những CHTN khó và những SV có trình độ năng lực cao được phân bố tiến dần lên phía trên (0,0), còn những CHTN dễ và những SV có khả năng thấp được phân bố tiến dần về phía dưới (0,0). Độ khó của các CHTN cùng khả năng của SV nằm trong khoảng từ (-3,0)

đến (4,0) theo đơn vị logic. Các CHTN số 17, 14, 8, 23.. là dễ nhất, còn các câu 31, 34, 21 là khó nhất. Ngoài việc so sánh các CHTN với nhau, nó cũng cho phép so sánh khả năng của lớp SV làm bài TN với độ khó của CHTN, có 12 SV có khả năng cao hơn độ khó của tất cả các CHTN . [6, 23].

Bảng số 4: Minh họa sự phù hợp của khả năng sinh viên làm bài TN theo Rasch

```
Run ONE:DE SO :16 TNSP DOT 2 10LOP K 05 . Test
-----
-----Case Fit In input Order
15/ 2/2008 8:42
all on viet (N = 108 L = 40)
-----
-----INFIT
MNSQ      .63      .71      .83      1.00      1.20      1.40
1.60
-----+-----+-----+-----+-----+-----+-----
--+-
  1 1A  T-C      .      |      *      .
  2 2A  T-C      .      | *      .
  3 3A  T-Q      .      | *      .
  4 4A  N-L      .      | *      .
  5 5A  H-D      .      *      .
  6 6A  H-D      .      *      |      .
  7 7A  N-D      .      *      |      .

39 39D D-A      .      *      |      .

106 106J L-K      .      |      .
*
107 107J N-T      .      |      *      .
108 108J L-T      .      *      |      .
```

Qua biểu đồ Casse fit In Infut order cho phép ta nhanh chóng xác định được những SV nào có những câu trả lời bất thường, tức là những câu trả lời không phù hợp với mô hình Rasch. Đây là những SV làm đúng một số câu vượt quá khả năng (Harder Achieved) hoặc đã làm sai một số CHTN dưới mức khả năng của mình (Easier not Achieved) trong bảng số 4, SV có số thứ tự 106J và được minh họa ở (Bảng số 5) biểu đồ Kidmap sau.

Bảng số 5: Phân bố thống kê tương qua giữa độ khó và kết quả làm bài của SV 106J

```
Run ONE:DE SO :16 TNSP DOT 2 10LOP K 05 . Test
```

----- K I D M A P -----			
Candidate: 106J L-K-TRONG		ability: -1.22	
group:	all	fit:	1.53
scale:	all	% score:	27.03
-----Harder Achieved -----		-----Harder Not Achieved -----	
(vùng trả lời đúng ngoài dự kiến)		(vùng trả lời sai trong dự kiến)	
		31 (C)	
34		21 (A)	
	32	12 (E)	
		28 (A)	40 (D)
		5 (E)	
26		15 (B)	
16		33 (E)	36 (C) 38 (A)
6		9 (A)	29 (B)
24		18 (C)	
		10 (D)	27 (B) 39 (E)
		7 (E)	
	25 13	2 (C)	
	1	35 (E)	
.....		3 (D)	11 (D)
(vùng trả lời đúng trong dự kiến)		20 (A)	
		19 (D)	22 (E)
		XXX	
	8	23 (A)	
		(vùng trả lời sai ngoài dự kiến)	
		14 (D)	
		17 (E)	
-----Easier Achieved -----		-----Easier Not Achieved -----	
=====			

Đối với SV có mã số 106J có một câu trả lời đúng trong dự đoán, còn 9 câu trả lời đúng ngoài dự đoán. Để tìm hiểu tình hình học tập rèn luyện của SV này ta dựa vào bảng Kidmap kết hợp với trao đổi trực tiếp, SV này thừa nhận học không kịp bài, để đối phó với đợt kiểm tra em đã copy và đoán mò. Do đó, sau các bài kiểm tra, dựa vào biểu đồ Case fit in infut order và Kidmap, giảng viên xem xét bài làm của các SV cá biệt để tìm hiểu điều gì có thể là nguyên nhân của các câu trả lời bất thường này.

5/ Các bảng số liệu trên minh họa kết quả phân tích 40 CHTN của một bài kiểm tra có 108 SV tham gia, chứng tỏ phần mềm Quest có thể được sử dụng rất tiện lợi và có hiệu quả trong việc phân tích số liệu các CHTN và đánh giá kết quả học tập của SV theo lý thuyết trắc nghiệm hiện đại IRT.

Vậy bằng phương pháp này cùng các biểu đồ trên tỏ ra là một phương tiện dạy học hiện đại không những giúp ta phân tích, chọn được những CHTN đạt yêu cầu lưu vào ngân hàng CHTN mà còn giúp cho giảng viên chẩn đoán thăm dò được tình hình học tập của các SV cá biệt, qua đó kịp thời giúp đỡ để các SV này tự điều chỉnh quá trình học của mình.

Chúng tôi cho rằng với ngân hàng CHTN đủ lớn, được soạn thảo và phân tích cẩn thận bao trùm toàn chương trình, các bài kiểm tra luân thay đổi nội dung sẽ có tác dụng tốt đến quá trình tự học, làm giảm sự may rủi, học tủ, học lệch, đồng thời hạn chế được nạn quay cóp, buộc SV phải lo học từ đầu năm. Nó có tác dụng định hướng học tập cho SV, tạo động cơ học tập tích cực, tự lực cho SV. Nó sẽ là một phương tiện để đổi mới phương pháp dạy, phương pháp học và quản lý chất lượng đào tạo ở đại học.

TÀI LIỆU THAM KHẢO

- [1] Đào Hữu Hồ (2000), *Thống kê xã hội học* - NXB Hà Nội, Hà nội
- [2] Donald E. Morrison (1993), *Applied Linear Statistical Methods*, Prentice Hall.
- [3] Frederick J.G, Larry B.W (1992) *Statistic for the Behavioral Sciences*. Third Edition. New York .West Publing company
- [4] Parick Griffin (1994), *Testing and evaluation*. Tài liệu dùng cho lớp tập huấn tại HCM - Huế - Hà Nội
- [5] Raymond.J. Adams, Sick Toan Khoa. *Quest the interactive Test Analysis System – ACER*.
- [6] Dương Thiệu Tống (1998), *Trắc nghiệm tiêu chí* - NXBGD Hà Nội
- [7] Dương Thiệu Tống (1995), *Trắc nghiệm và đo lường thành quả học tập*, ĐHTH Tp. HCM