

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

Nguyễn Thị Ngọc Xuân *

Tóm tắt

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

Từ khóa: Phần mềm Quest/Conquest, Anh văn không chuyên, lý thuyết đáp ứng câu hỏi.

Abstract

The paper presents the results of using Quest/Conquest software, which is used to analyze a test through multiple-choice method. Quest/Conquest is software for analyzing and evaluating multiple choice questions as well as tasks based on Item Response Theory (IRT).

Key Words: Quest/Conquest software, non-professional English, item response theory.

1. Đặt vấn đề

Trắc nghiệm là một phương pháp của khoa học về đo lường trong giáo dục. Những năm gần đây, trắc nghiệm khách quan (TNKQ) được đưa vào trong các kỳ thi tốt nghiệp phổ thông và tuyển sinh đại học nên sự quan tâm của các trường học Đại học Sư phạm, các cơ sở quản lý giáo dục và đội ngũ giáo viên về phương pháp giảng dạy này ngày một tăng lên. Một ưu điểm nổi bật của phương pháp trắc nghiệm là việc sử dụng các phần mềm để xử lý số liệu của câu hỏi trắc nghiệm (CHTN), đề trắc nghiệm (ĐTN) đã mang lại hiệu quả cao. Do đó, việc vận dụng phần mềm Quest/Conquest để phân tích, đánh giá câu hỏi trắc nghiệm, đề trắc nghiệm là rất cần thiết.

2. Giới thiệu về lý thuyết Ứng đáp câu hỏi

Thuyết ứng đáp câu hỏi (Item Response Theory - IRT) là một lý thuyết của khoa học về đo lường trong giáo dục, ra đời từ nửa sau của thế kỷ XX và phát triển mạnh mẽ cho đến nay. So với lý thuyết khảo thí cổ điển, lý thuyết khảo thí hiện đại ưu việt hơn nhiều, được áp dụng ngày càng rộng rãi để định cỡ các CHTN và thiết kế các đề trắc nghiệm.

Thuyết ứng đáp câu hỏi của Rasch mô hình hóa mối quan hệ giữa mức độ khả năng của người làm trắc nghiệm và đáp ứng của người ấy với câu trắc nghiệm. Mỗi câu trắc nghiệm được mô tả bằng một thông số (độ khó) ký hiệu là δ và mỗi người làm trắc nghiệm được mô tả bằng một thông số (khả năng) ký hiệu là θ . 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:

$$P(\theta) = \frac{\exp(\theta - \delta)}{1 + \exp(\theta - \delta)}$$

Trong đó, $P(\theta)$ là xác suất để thí sinh n có năng lực θ trả lời ĐÚNG câu hỏi có độ khó δ .

3. Xử lý số liệu bằng phần mềm Quest/Conquest

Chúng tôi sử dụng mô hình Rasch với phần mềm Quest/Conquest để phân tích đề thi trắc nghiệm khách quan môn tiếng Anh của sinh viên (SV) Trường Đại học Trà Vinh gồm 60 câu hỏi trắc nghiệm. Đề thi có 1150 thí sinh tham gia với 5 mã đề khác nhau (có chất lượng tương đương nhau). Chúng tôi đã chọn ngẫu nhiên 1 mã đề bất kỳ trong 5 mã đề.

- Sau khi thi xong, Phòng Khảo thí và Đảm bảo Chất lượng của Trường tiến hành thu nhận bài thi. Kết quả bài thi được nhập bằng tay vào máy tính. Dữ liệu này được đưa vào phần mềm Quest/Conquest.

- Chương trình Quest/Conquest xử lý sẽ cho ta các kết quả như mức độ phù hợp của các câu hỏi với mô hình Rasch, năng lực của thí sinh so với độ khó của câu hỏi, độ tin cậy của đề thi và các chỉ số đặc trưng cho từng câu hỏi như độ khó, độ phân biệt, hệ số tương quan giữa câu hỏi thi với toàn bài, độ tin cậy, sai số.

- Sau đây là các bảng mô tả kết quả phân tích 60 câu trắc nghiệm.

3.1. Mức độ phù hợp với mô hình

Khi dữ liệu phù hợp với mô hình Rasch thì trị số kỳ vọng của các bình phương trung bình (mean square) xấp xỉ bằng 1 và độ lệch chuẩn SD xấp xỉ bằng 0.

Bảng 1: Mức độ phù hợp với mô hình Rasch

```

Summary of item Estimates
=====
Mean
SD
SD (adjusted)
Reliability of estimate
Fit Statistics
=====
Infit Mean Square      Outfit Mean Square
Mean      1.00          Mean      1.00
SD         .02             SD         .03
Summary of case Estimates
=====
Mean
SD
SD (adjusted)
Reliability of estimate
Fit Statistics
=====
Infit Mean Square      Outfit Mean Square
Mean      1.00          Mean      1.00
SD         .06             SD         .08
Infit t          Outfit t
Mean      .03             Mean      .05
SD         .76             SD         .45
    
```

- Từ các số liệu về giá trị trung bình Mean và độ lệch chuẩn SD có được từ file xuan.map khi xử lý dữ liệu bằng phần mềm QUEST, kết quả cho thấy dữ liệu dùng để phân tích là phù hợp với mô hình Rasch.

- Các thông tin về kết quả tính toán cho thấy năng lực trung bình của mẫu thí sinh (case estimate) tham gia bài kiểm tra 0,45 hơn nhiều so với độ khó chung của bài kiểm tra. Độ tin cậy của tính toán rất đáng tin cậy vì có giá trị bằng 0,93.

3.2. Mức độ phù hợp của các câu hỏi với nhau

Trong biểu đồ Item Fit sau đây, mỗi câu trắc nghiệm biểu thị bằng dấu *. Những câu trắc nghiệm nằm trong hai đường chấm thẳng đứng có giá trị INFIT MNSQ nằm trong khoảng [0,77; 1,30] sẽ phù hợp với mô hình Rasch. Nếu câu trắc nghiệm nào nằm ngoài khoảng này là không phù hợp và sẽ bị loại bỏ.

Bảng 2: Minh họa sự phù hợp các câu hỏi trong bài trắc nghiệm

```

DE THI TINHOCDAI CUONG
-----
Item Fit
17/ 6/13 15:50
all on xuan (N = 228 L = 60 Probability
Level= .50)
-----
INFIT
MNSQ .63 .77 1.00 1.10 1.30 1.40
-----+-----+-----+-----+-----+-----+
1 item 1 . *
2 item 2 . | *
3 item 3 . * |
4 item 4 . *
5 item 5 . | *
.
.
.
59 item 59 . | *
60 item 60 . | *
-----
    
```

Kết quả cho thấy trong 60 câu hỏi nằm trong khoảng đồng bộ cho phép. Điều đó chứng tỏ 60 câu hỏi trong đề thi này đo đúng cái cần đo.

3.3. Phân bố độ khó câu hỏi thi và năng lực thí sinh

Sơ đồ phân bố độ khó câu hỏi thi và năng lực thí sinh cho thấy mức độ phù hợp của đề thi đối với thí sinh dự thi. Kết quả xử lý bằng phần mềm QUEST cho một bản đồ phân bố năng lực học sinh và độ khó câu hỏi thi.

Dựa vào biểu đồ ở bảng 2, ta thấy đề thi khá dễ với năng lực của nhóm thí sinh tham gia bài thi này, do đó cần bổ sung một số câu khó để đánh giá những thí sinh có năng lực cao, ngưỡng năng lực dưới -1. là 03, chưa có câu hỏi nào để đánh giá.

3.4. Độ tin cậy của đề thi: Kết quả tính toán bằng phần mềm Quest cho thấy độ tin cậy của đề thi đạt **0,93**. Đây là một đề thi có độ tin cậy cao.

3.5. Phân tích các tiêu chí khác

Ta tiếp tục xem xét các chỉ số thu được từ kết quả phân tích bằng phần mềm QUEST như sau:

- *Categories*: câu chọn, trắc nghiệm, phương án đúng được đánh dấu (*).

- *Disc*: độ phân biệt của câu hỏi giữa các nhóm thí sinh, (Disc) phải nằm trong khoảng 0,25 - 0,75 đối với các test trong lớp học.

- *Percent*: tỉ lệ phần trăm của một phương án là tỉ lệ giữa số thí sinh chọn phương án đó so với thí sinh làm bài kiểm tra; *Infit MNSQ* phải nằm trong khoảng 0,77 – 1,30.

- *Beserial*: hệ số tương quan point biserial. Cần loại bỏ những câu hỏi có mối tương quan thấp hoặc dưới 0 sẽ làm tăng độ tin cậy của bài kiểm tra.

- *StepLabel 1*: Giữa giá trị 0 và 1 có một bước, thí sinh thực hiện được bước này khi trả lời đúng câu hỏi.

- *Thresholds*: ngưỡng để vượt qua, thực chất là độ khó của câu trắc nghiệm.

- *Error*: sai số trong tính toán.

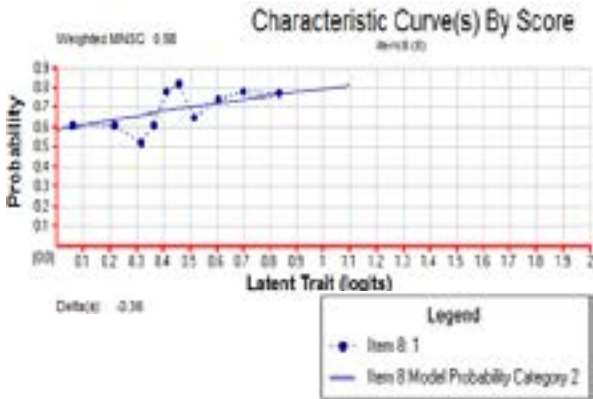
Bảng 4: Chỉ số thống kê của các câu hỏi thi được tạo ra từ QUEST

```

-----
Item 58:item 58          Infit MNSQ = .99
                          Disc = .19
Categories  A  B*  C  D  F  missing
Count       68  58  40  62  0
Percent(%)  29.8  25.4  17.5  27.2  .0
Pt-Biserial -.06 .19 -.03 -.10 NA
p-value     .188 .002 .305 .070 NA
Mean Ability .41 .56 .42 .39 NA NA
Step Labels          1
Thresholds          1.53
Error               .15
    
```

Ví dụ như câu 58 (item 58), có độ khó rất cao là 0,25, độ phân biệt rất thấp là 0,19, Infit MNSQ = 0,99 nằm trong khoảng cho phép, cho thấy đây là

Đồ thị biểu thị xác suất trả lời đúng câu hỏi với năng lực của thí sinh khi phân tích bằng Conquest ở Hình 3 dưới đây cũng cho thấy câu hỏi 8 là câu dễ, độ phân biệt không cao, đánh giá tốt đối với nhóm thí sinh có năng lực thấp.



Hình 3: Hình vẽ thể hiện xác suất trả lời đúng ở mức khá cao của câu 8 có độ khó dễ nhất (delta=-0,36).

Các câu còn lại dựa vào biểu đồ Item Estimates trình bày sự phân bố thống kê tương quan giữa năng lực của SV (dấu 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 30, 55,56,57,12,59,60 khó nhất, câu 17,9,4,5,6,29,10,20,48,11,27,22,52,14,37,28,47 là dễ nhất. Các thông tin về kết quả tính toán từ bảng Summary of case Estimates cho thấy năng

lực trung bình của mẫu thí sinh (case) tham gia bài kiểm tra (0.36) lớn hơn và gần bằng so với độ khó chung của bài kiểm tra (0.00) cho thấy độ khó của đề thi tương đối dễ so với năng lực thí sinh và được thể hiện rõ ràng hơn ở biểu đồ mô tả mối quan hệ giữa năng lực của thí sinh và độ khó của câu hỏi.

Điều này cho thấy bài test này quá dễ so với năng lực của thí sinh, do đó cần tăng số lượng các câu hỏi khó để đánh giá những học sinh có năng lực cao, cần điều chỉnh các câu 2, 14, 22, 27, 48, 54, 56, 60 (đặc biệt là câu 14) để đề thi tốt hơn.

4. Kết luận

Các bảng số liệu trên minh họa kết quả phân tích 60 CHTN chứng tỏ phần mềm Quest/Conquest 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 khảo thí hiện đại.

Như vậy, nhờ có sự hỗ trợ của hai phần mềm chuyên dụng Quest/Conquest, chúng ta đã phân tích đề thi một cách nhanh chóng, tiện lợi và có được cái nhìn toàn diện về kết quả như sau: Chất lượng đề thi tương đối tốt, đề thi khá dễ đối với thí sinh kiểm tra ít câu hỏi khó, do đó cần phải xem xét các câu hỏi này cho phù hợp với mức độ của kỳ thi.

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.

Tài liệu tham khảo

Phạm Xuân Thanh. 2013. *Bài giảng môn Lý thuyết đo lường và đánh giá*.
 Phạm Xuân Thanh. 2013. *Bài giảng môn Mô hình Rasch và Phân tích dữ liệu bằng phần mềm QUEST*.
 Phạm Xuân Thanh. 2008. *Tiểu đề án Phân tích câu hỏi thi của các đề thi trắc nghiệm khách quan*.
 Phạm Xuân Thanh. 2005. *Slide tập huấn Phân tích kết quả điều tra*.
 Lâm Quang Thiệp. 2011. *Đo lường trong giáo dục, lý thuyết và ứng dụng*. Nhà xuất bản Đại học Quốc gia Hà Nội.
 Lâm Quang Thiệp, Lâm Ngọc Minh, Lê Mạnh Tấn, Vũ Đình Bông. 2007. *Phần mềm Vitesta và việc phân tích số liệu trắc nghiệm*. Tạp chí giáo dục. Số 176. 11/2007.
 Nguyễn Bảo Hoàng Thanh. 2008. *Sử dụng phần mềm Quest để phân tích câu hỏi trắc nghiệm khách quan*. Tạp chí Khoa học và Công nghệ. Đại học Đà Nẵng. số 2(25)2008.
 Đặng Thị Hương. 2012. *Xây dựng hệ thống câu hỏi trắc nghiệm khách quan bằng phần mềm Vitesta*. Tuyển tập Báo cáo Hội nghị Sinh viên Nghiên cứu Khoa học lần thứ 8 Đại học Đà Nẵng.