VJE Tạp chí Giáo dục, Số 488 (Kì 2 - 10/2020), tr 12-17 ISSN: 2354-0753

SỬ DỤNG PHẦN MỀM WINSTEPS ĐỂ PHÂN TÍCH CÂU HỎI BÀI THI TỰ LUẬN

Nguyễn Thị Ngọc Xuân[1,+], Nguyễn Thành Chiến[2]

[1]Trường Đại học Trà Vinh; [2]Trường Đại học Quy Nhơn
+Tác giả liên hệ ● Email: ngocxuan@tv.edu.vn

Article History
Received: 25/8/2020
Accepted: 07/9/2020
Published: 20/10/2020

Keywords
Winsteps software, analysis, tests, essay.

ABSTRACT
The essay format plays an important role in evaluating a student's learning outcomes. However, there are not many research projects using Winsteps software to analyze and evaluate the quality of essay test in order to detect errors, thereby promptly correct and supplement to improve the quality of exam questions. This article presents the analysis results of the essay exam for the "object-oriented programming" module at Binh Dinh College with Winsteps software built on the basis of theoretical model of IRT (Item Response Theory). The analysis results with this software give us the specific, accurate, fast and intuitive numbers about exam characteristics, exam questions and competitor's competencies as well as the relationship between the factors. This is in the form of text and graph. This is one of the important bases for designing evaluation tool to have the basis to correct or eliminate unsatisfactory questions.

1. Mở đầu

Hiện nay, chất lượng giáo dục đang trở thành mối quan tâm chung của toàn xã hội, đặc biệt là chất lượng giáo dục đại học trong bối cảnh đổi mới giáo dục theo hướng tiếp cận năng lực của người học. Một trong những yêu cầu cấp thiết nhằm nâng cao chất lượng giáo dục đại học chính là việc đổi mới phương pháp giảng dạy, trong đó có đổi mới về phương pháp kiểm tra, đánh giá kết quả học tập một cách toàn diện và đúng năng lực của người học. Có thể nói, việc kiểm tra, đánh giá là hoạt động không thể thiếu của quá trình dạy và học. Thông qua kiểm tra, đánh giá trình độ nhận thức, kĩ năng, kĩ xảo của người học sẽ phát hiện những sai sót, những lỗ hổng về kiến thức…, từ đó giúp người dạy và người học tự điều chỉnh hoạt động dạy học đạt chất lượng tốt hơn.

Bên cạnh hình thức thi trắc nghiệm đang được các cơ sở giáo dục áp dụng và đạt được một số lợi ích nhất định thì hình thức thi tự luận vẫn còn giữ một vai trò quan trọng trong đánh giá kết quả học tập của học sinh. Tuy nhiên, ở nước ta chưa có nhiều công trình nghiên cứu sử dụng mềm Winsteps để phân tích đánh giá chất lượng các đề thi tự luận (essay test) bằng các phần mềm chuyên dụng nhằm phát hiện những sai sót, từ đó kịp thời chỉnh sửa bổ sung nâng cao chất lượng đề thi. Bài viết này dựa trên mô hình lí thuyết hồi đáp IRT (Item Response Theory) và sử dụng phần mềm Winsteps để phân tích câu hỏi bài thi tự luận học phần "lập trình hướng đối tượng" tại Trường Cao đẳng Bình Định.

2. Kết quả nghiên cứu

2.1. Lí thuyết hồi đáp

Từ sau thập niên 70 của thế kỉ XX đến nay, lí thuyết phân tích câu hỏi hiện đại ra đời dựa trên mô hình toán học, đòi hỏi nhiều tính toán, nhưng nhờ sự tiến bộ của công nghệ tính toán bằng máy tính điện tử trong mấy thập niên qua nên đã phát triển nhanh chóng và đạt được những thành tựu quan trọng. Cốt lõi của lí thuyết khảo thí hiện đại là "Lí thuyết đáp ứng câu hỏi" (Item Response Theory - IRT) (Lâm Quang Thiệp, 2012). IRT dựa trên hai giả thuyết: (1) Sự đáp ứng của một thí sinh đối với một câu hỏi có thể được tiên đoán bằng năng lực tiềm ẩn của thí sinh; (2) Quan hệ giữa sự đáp ứng câu hỏi của thí sinh và năng lực tiềm ẩn làm cơ sở cho sự đáp ứng đó có thể mô tả bằng một hàm "Đặc trưng câu hỏi" đồng biến.

Để xây dựng một mô hình toán diễn tả một mối quan hệ phải xuất phát từ một tiền đề nào đó. Nhà toán học George Rasch (người Đan Mạch) đã xây dựng được một mô hình hàm đặc trưng câu hỏi đơn giản nhất nhưng đến nay vẫn được sử dụng nhiều nhất trong công nghệ đo lường giáo dục (Lâm Quang Thiệp, 2008).

Thuyết đáp ứng câu hỏi của Rasch mô hình hóa mối quan hệ giữa mức độ, khả năng của người làm câu hỏi và đáp ứng của người ấy với câu hỏi. Mỗi câu hỏi được mô tả bằng một thông số (độ khó) kí hiệu là δ và mỗi người làm câu hỏi được mô tả bằng một thông số (khả năng) kí 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 câu hỏi ấy (Phạm Xuân Thanh, 2013). Dạng toán học của mô hình này là:

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

Trong đó, P(θ) là xác suất để thí sinh *n* có năng lực θ trả lời ĐÚNG câu hỏi có độ khó δ. Xác suất trả lời đúng một câu hỏi được biểu thị bằng một hàm số liên quan đến năng lực của thí sinh θ (thể hiện dưới một hàm số logarit tự nhiên). Xác suất đó là hàm P(θ) - một đường cong đặc trưng của câu hỏi (Item Characteristic Curves - ICC). Hiện nay, với sự phát triển mạnh mẽ của công nghệ thông tin, ta có thể dễ dàng tính được năng lực θ và độ khó δ bằng các phần mềm chuyên dụng như Quest, Conquest, Winsteps…

2.2. Xử lí bằng phần mềm Winsteps

Winsteps là phần mềm xử lí dữ liệu chuyên dụng được xây dựng trên môi trường hệ điều hành Windows, hỗ trợ nhiều ứng dụng của mô hình Rasch đặc biệt trong các lĩnh vực đo lường đánh giá trong giáo dục, khảo sát thái độ (attitude surveys) và phân tích thang đánh giá (rating scale analysis) (John, 2020). Winsteps được xây dựng từ kết quả nghiên cứu "Rating Scale Analysis" (Phân tích thang đánh giá) của tác giả Benjamin D. Wright, Geofferey N. Masters tại Trường Đại học Chicago năm 1982. Hiện nay, có rất nhiều nhà khoa học trên khắp thế giới vẫn sử dụng phần mềm Winsteps như là một công cụ chính trong việc xử lí dữ liệu của nhiều lĩnh vực khác nhau: Tâm lí học đường (William, 2017), Y học lâm sàng (Velozo et al., 2009)… Winsteps được đánh giá là một trong những phần mềm sử dụng rộng rãi nhất vì giao diện thân thiện người dùng và được tác giả của phần mềm cung cấp hướng dẫn chi tiết, trợ giúp cho người sử dụng (William, 2016).

Quá trình xử lí dữ liệu bằng phần mềm Winsteps được thực hiện như sau: sau khi thi xong, dữ liệu bài thi của 150 sinh viên được nhập vào phần mềm SPSS xử lí số liệu thô (mỗi dòng tương ứng với kết quả của một thí sinh, mỗi cột tương ứng với mỗi ý trong đề thi tự luận, đề thi gồm có 3 câu chia thành 23 ý nhỏ). Tiếp theo đó, sử dụng phần mềm Winsteps phân tích chất lượng câu hỏi 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; các tham 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. Dựa kết quả phân tích này, các câu hỏi không đạt chất lượng (không phù hợp với mô hình, các tham số không đạt yêu cầu) sẽ được rà soát điều chỉnh hoặc loại bỏ để góp phần nâng cao chất lượng các câu hỏi.

2.3. Phân tích câu hỏi

2.3.1. Phân tích mức độ phù hợp với mô hình

Khi dữ liệu phù hợp với mô hình Rasch thì trị số kì vọng của các bình phương trung bình (mean square) xấp xỉ bằng 1 và độ lệch chuẩn (SD) xấp xỉ bằng 0. Qua kết quả thu được, ta có ***MNSQ INFIT=1.03, SD INFIT=0.27, MNSQ OUTFIT=1.01, SD OUTFIT=0.63*** chứng tỏ kết quả tính toán phù hợp với mô hình Rasch.

2.3.2. Độ tin cậy của đề thi

Kết quả tính toán bằng phần mềm Winsteps cho thấy độ tin cậy Cronbach Alpha (KR-20) của đề thi bằng **0.85** và mối tương quan giữa năng lực và điểm thô bằng **0.96**. Qua đó chứng tỏ đây là một đề thi có độ tin cậy cao và kết quả đo được phản ánh đúng năng lực thật của thí sinh.

PERSON RAW SCORE-TO-MEASURE CORRELATION = .96

CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .85

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

Trong quá trình xử lí, Winsteps còn tạo ra biểu đồ xử lí Item Fit, mỗi câu hỏi biểu thị bằng dấu *. Những câu hỏi nằm trong hai đường chấm thẳng đứng có giá trị trung bình bình phương độ phù hợp INFIT MNSQ nằm trong khoảng (0.77; 1.30) sẽ phù hợp với mô hình Rasch. Nếu câu hỏi nào không phù hợp thì loại bỏ.

Bảng 1. Minh hoạ sự phù hợp của một số câu hỏi trong đề thi

ITEM FIT GRAPH: MISFIT ORDER

ENTRY NUMBER	MEASURE - +	INFIT MEAN-SQUARE 0 0.7 1 1.3 2	OUTFIT MEAN-SQUARE 0 0.7 1 1.3 2	ITEM G
11	*	: . : *	A : . : *	Cau2a1 0
12	*	: . :*	B : . : *	Cau2a2 0
19	*	: . :*	C : . : *	Cau2b7 0
1	*	: .* :	D : . :*	Cau1a1 0
2	*	: .* :	E : . :*	Cau1a2 0
3	*	: .* :	F : . :*	Cau1a3 0

```
|     16|       * |       :  . *:       |G      :  .* :       | Cau2b4 0 |
|     15|      *  |       :  . *:       |H      :  .* :       | Cau2b3 0 |
|     10|    *    |       :  .* :       |I      : *.  :       | Cau1b6 0 |
|      4|  *      |       :  *  :       |J      :  *  :       | Cau1a4 0 |
|     20|      *  |       : *.  :       |K      *  .  :       | Cau3a  0 |
|     18|      *  |       : *.  :       |k      : *.  :       | Cau2b6 0 |
|      5| *       |       : *.  :       |j     *:  .  :       | Cau1b1 0 |
|     13|     *   |       :* .  :       |i      *  .  :       | Cau2b1 0 |
|      8|   *     |       :* .  :       |h    * :  .  :       | Cau1b4 0 |
|      7|*        |       :* .  :       |g*     :  .  :       | Cau1b3 0 |
|      9|   *     |       *  .  :       |f   *  :  .  :       | Cau1b5 0 |
|     14|      *  |       *  .  :       |e      *  .  :       | Cau2b2 0 |
|     23|        *|       *  .  :       |d   *  :  .  :       | Cau3d  0 |
|     17|     *   |       *  .  :       |c   *  :  .  :       | Cau2b5 0 |
|     22|       * |       *  .  :       |b   *  :  .  :       | Cau3c  0 |
|     21|      *  |       *  .  :       |a    * :  .  :       | Cau3b  0 |
-----------------------------------------------------------------------------
```

Trong *bảng 1* ở trên, ta thấy có 3 câu (gồm câu 11, 12, 19) là biến ngoại lai, có chỉ số (INFIT MNSQ) nằm ngoài vùng cho phép, cần loại bỏ. Còn các câu khác đều đạt yêu cầu vì nằm trong khoảng chấp nhận của mô hình.

2.3.4. Phân tích 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. Theo kết quả phân tích ta có độ khó trung bình của các câu hỏi bằng 0 nhỏ hơn rất nhiều so với năng lực trung bình của thí sinh là 1.85 chứng tỏ đây là một đề thi khá dễ so với nhóm thí sinh tham gia đánh giá. Với đồ thị phân bố năng lực học sinh và độ khó câu hỏi thi được tạo ra từ phần mềm Winsteps cho ta thấy được kết luận này khá rõ ràng.

Bảng 2. Phân bố năng lực thí sinh và độ khó câu hỏi thi (thang logistic)

```
MEASURE PERSON - MAP - ITEM
         <more>|<rare>
     4  XXXXXX  +
                |T
NĂNG LỰC CAO   T|            RẤT KHÓ
                |
              X |
                |
          XXXX  |
     3          +  Cau2a2
          XXXX  |
           XXX S|
          XXXX  |  Cau3d
           XXX  |
         XXXXX  |
             X  |  Cau3c
     2     XXX  +  Cau2b4
          XXXX  |S
            XX  |  Cau3b
       XXXXXXX M|
           XXX  |  Cau2b2  Cau2b3  Cau2b7
                |
         XXXXX  |  Cau2b6
     1     XXX  +  Cau3a
            XX  |
            XX  |
           XXX  |
             X S|  Cau2b5
            XX  |  Cau2b1
                |
     0       X  +M
```

Kĩ năng xây dựng các phương thức có thuật toán phức tạp

Kĩ năng xây dựng các phương thức cơ bản trong xây dựng lớp (hướng đối tượng)

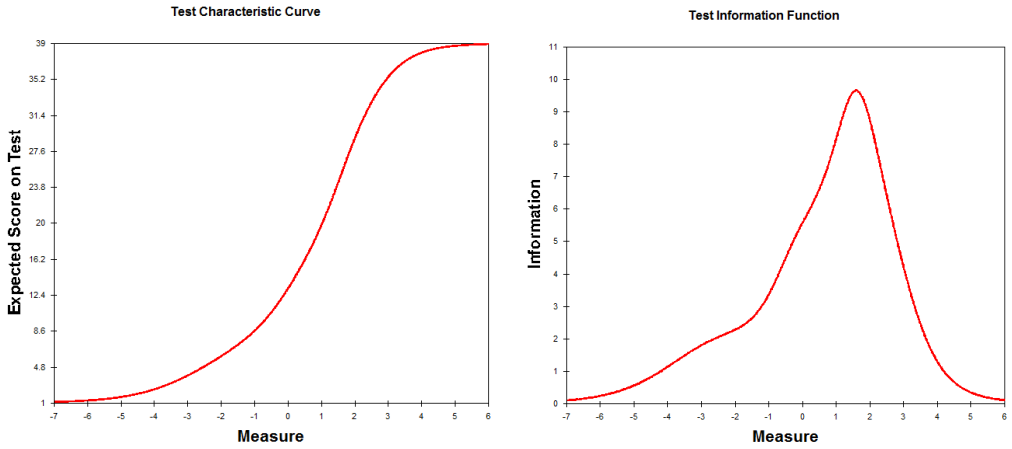
```
                    |  Cau1b6
                    |  Cau2a1
              XXX   |
                X T |
                    |
                    |
                    |
      -1            +
                    |  Cau1b5
                    |  Cau1b4
                    |
               XX   |
                    |
                    |S
      -2            +
                    |  Cau1a4
                    |  Cau1a1   Cau1a2   Cau1a3   Cau1b1
                    |
      -3            +
                    |
                    |  Cau1b3
NĂNG LỰC THẤP       |          RẤT DỄ
                    |
                    |
                    |T
      -4            +
            <less>  |<frequent>
   Mỗi dấu x đại diện cho 1 sinh viên
================================================================================
```

Kĩ năng khởi tạo lớp, đối tượng

Bảng 2 cho thấy, đề thi 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 trên 3, hiện tại đề thi còn khá ít câu hỏi đánh giá ở mức này.

Với kết quả phân tích các câu hỏi ở trên, có thể thấy câu hỏi đề thi phân thành 3 nhóm chính: *nhóm 1* gồm các câu hỏi dễ nhất với *ngưỡng thresholds dưới -1* đánh giá kĩ năng khởi tạo lớp, đối tượng mới trong lập trình hướng đối tượng; *nhóm 2* gồm các câu hỏi ở mức trung bình với *ngưỡng thresholds từ -1 đến 1* đánh giá kĩ năng xây dựng các phương thức cơ bản trong xây dựng lớp; *nhóm 3* gồm các câu hỏi ở mức năng lực từ với *ngưỡng thresholds từ 1 trở lên* đánh giá kĩ năng xây dựng các phương thức có thuật toán phức tạp.

Ngoài ra, hình dáng của đường cong đặc trưng và điểm cực đại của đường cong thông tin của đề thi ở *hình 1* cho thấy đề thi là dễ và đánh giá chính xác năng lực của những thí sinh ở mức trung bình, khá.



Hình 1. Đường cong đặc trưng đặc trưng và thông tin của cả bài thi

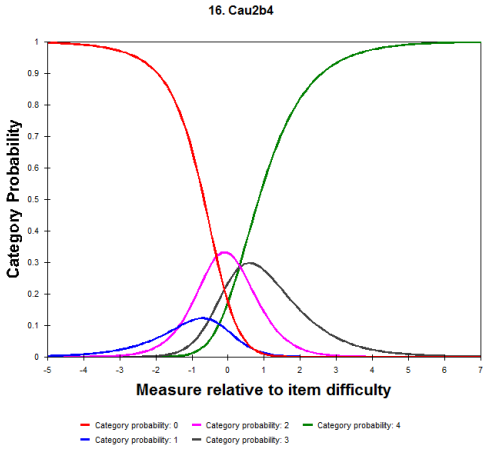
Đánh giá câu hỏi thi qua một số chỉ số (bảng 3):

Bảng 3. Thông tin đặc trưng của câu hỏi

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE ABILITY	S.E. MEAN	OUTF MNSQ	PTMEA CORR.	ITEM
18	0	0	21	28	.50	.22	.6	-.61	Cau2b6
	1	1	8	11	1.90	.39	2.2	.01	
	2	2	46	61	2.46	.17	.9	.55	

Qua kết quả thu được từ số liệu phân tích ở *bảng 3*, chúng ta còn thu được một số thông tin quan trọng về câu hỏi thi như các mức cho điểm (Score value), tỉ lệ trả lời (Data Count), năng lực trung bình trả lời câu hỏi (Average Ability), chỉ số SE MEAN, OUTF MNSQ, mối tương quan của ý câu hỏi với điểm toàn bài. Dựa vào các chỉ số này ta có thể dự đoán sự hợp lí chất lượng câu hỏi tương ứng ở các mức điểm. Ví dụ: câu 2b6 dưới đây cần xem xét lại mức 1 vì có mối tương quan khá thấp và chỉ số OUTFIT MNSQ cũng vượt ra ngoài vùng (0,77:1.3)

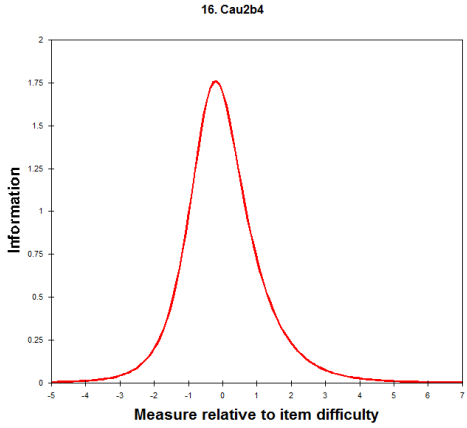
Đánh giá câu hỏi qua đồ thị dưới đây (hình 2):



Hình 2. Biểu diễn mối tương quan xác suất trả lời đúng câu hỏi ý 2b4 và độ khó câu hỏi

Biểu đồ ở *hình 2* cho thấy, câu 2b4 (số thứ tự 16) là một câu hỏi khó với 4 mức điểm trả lời, các thí sinh ở mức năng lực dưới -1 thì chỉ làm được tối đa ở mức 2, thí sinh có từ 0 đến 1 chỉ làm được mức 3 và *chỉ các thí sinh có năng lực từ 1 trở lên mới có xu hướng trả lời được điểm tối đa của câu hỏi này.*

Tuy nhiên, câu hỏi này lại đánh giá tốt đối với nhóm thí sinh có năng lực trung bình và khá thể hiện qua *điểm cực đại* của đồ thị ở *hình 3*.



Hình 3. Đường cong đặc trưng thông tin của câu hỏi 2b4

2.4. Phân tích năng lực thí sinh

Theo số liệu phân tích thì năng lực của thí sinh nằm trong khoảng -1.55 đến 4.95 và có 6 thí sinh có mức năng lực cao nhất và ngang bằng nhau. Năng lực trung bình của nhóm thí sinh này là 1.85 với độ lệch chuẩn là 1.39 được chia thành 4 nhóm tương ứng với 4 mức năng lực đã phân tích ở trên.

Đề thi có 8 thí sinh với outfit MNSQ vượt quá 1.3 và 46 thí sinh (chiếm 61%) với outfit MNSQ thấp hơn 0.77, do đó đề thi này có nhiều thí sinh trả lời đúng các câu hỏi mà độ khó cao hơn so với năng lực thật sự của họ.

Bảng 4. Thống kê các thí sinh có câu trả lời không mong đợi

```
   MOST MISFITTING RESPONSE STRINGS
        PERSON OUTMNSQ  |ITEM
        |        1111211112121
        |753214891037085491632
       high--------------------
40 43     9.90 A|..0000...0.........2.
34 37     3.96 B|........0....0.......
12 12     3.95 C|........0............
61 64     1.03 D|................0.0.2
```

Ngoài ra, phần mềm Winsteps có thể phát hiện những câu trả lời không mong đợi (câu trả lời ngoài sự mong muốn của người thiết kế câu hỏi) từ phía thí sinh và biểu diễn một cách rất rõ ràng. Như kết quả ở *bảng 4* cho thấy, thí sinh có số thứ tự (STT) 40 có đến 6 câu trả lời không mong đợi, thí sinh STT 34 có 2 câu trả lời không mong đợi, thí sinh STT 34 và 12 có cùng câu trả lời sai với cùng một nội dung câu hỏi.

3. Kết luận

Như vậy, nhờ có sự hỗ trợ của phần mềm chuyên dụng WINSTEPS, chúng ta đã phân tích đề thi một cách nhanh chóng, tiện lợi và có được góc nhìn toàn diện về kết quả đánh giá như sau: chất lượng đề thi tương đối tốt, đề thi hơi dễ so với nhóm thí sinh tham gia kiểm tra, đánh giá. Độ tin cậy của đề thi ở mức rất cao; tuy nhiên, cần xem xét một số câu đánh giá không đúng so với năng lực của thí sinh gây ra những câu trả lời không mong đợi.

Bằng cách thức phân tích kết quả số liệu kết hợp với việc sử dụng biểu đồ trực quan như trên xem ra là một phương pháp hiện đại không những giúp ta phân tích, chọn lọc được những câu hỏi đạt yêu cầu đưa vào ngân hàng đề thi mà còn hỗ trợ cho giảng viên chẩn đoán thăm dò được nguyên nhân sai sót của các câu hỏi kém chất lượng từ đó có biện pháp điều chỉnh, bổ sung cho phù hợp cho các đợt kiểm tra tiếp theo.

Tài liệu tham khảo

David Andrich (1998). *Rasch models for measurement* - SAGE Publication, USA.

Dương Thiệu Tống (2005). *Trắc nghiệm và đo lường thành quả học tập (phương pháp thực hành)*. NXB Khoa học xã hội.

Griffin, P. (1998). *An Introduction to the Rasch Model. (Draft). Melbourne: Assessment Research Cente*r. Melbourne University, Vic.

Griffin, P. (2000). *Program Development and Evaluation.* Assessment Research Centre, the University of Melbourne.

John M.Linacre (2020). *A User's Guide to Winteps Ministep Rasch Model Computer Programs.* Chicago.

Lâm Quang Thiệp (2008). *Trắc nghiệm và Ứng dụng*. NXB Khoa học và Kĩ thuật.

Lâm Quang Thiệp (2012). *Đo lường và đánh giá hoạt động học tập trong nhà trường*. NXB Đại học Sư phạm.

Nguyễn Thị Ngọc Xuân (2014). *Sử dụng phần mềm Quest/ConQuest phân tích câu hỏi trắc nghiệp khách quan.* Tạp chí Khoa học, Trường Đại học Trà Vinh, số tháng 9, tr 24-27.

Phạm Xuân Thanh (2013). *Bài giảng môn Mô hình Rasch và Phân tích dữ liệu bằng phần mềm QUEST*. Hà Nội.

Velozo CA, Ying Wang, Leigh Lehman, Jia-Hwa Wang (2009). *Utilizing Rasch measurement models to develop a computer adaptive self-report of walking, climbing, and running*. Disability and Rehabilitation. DOI: https://doi.org/10.1080/09638280701617317.

William J. Boone (2016). *Rasch Analysis for Instrument Development: Why, When, and How*? CBE: Life Sciences Education. DOI: https://doi.org/10.1187/cbe.16-04-0148.

William J. Boone, Amity Noltemeyer (2017). *Rasch analysis: A primer for school psychology researchers and practitioners.* Cogent Education. DOI: https://doi.org/10.1080/2331186X.2017.1416898.