

## SỬ DỤNG BẢNG GSP VÀ PHƯƠNG PHÁP ROC ĐỂ PHÂN TÍCH VÀ LỰA CHỌN CÂU HỎI TRẮC NGHIỆM KHÁCH QUAN

• Nguyễn Phước Hải<sup>(\*)</sup>

### Tóm tắt

*Bài viết đề xuất phương pháp phân tích và lựa chọn câu hỏi trắc nghiệm khách quan dựa trên bảng GSP và phương pháp ROC với trường hợp cỡ mẫu lớn. Trong nghiên cứu này, phương pháp đề xuất đã được so sánh với lý thuyết ứng đáp câu hỏi. Kết quả nghiên cứu đã cho thấy phương pháp này không chỉ có thể sử dụng phân tích và lựa chọn câu hỏi trắc nghiệm khách quan, mà còn có thể cải thiện chất lượng và hiệu quả của việc thiết kế câu hỏi để xây dựng ngân hàng câu hỏi trắc nghiệm khách quan.*

*Từ khóa: bảng GSP, phương pháp ROC, lý thuyết ứng đáp câu hỏi, câu hỏi trắc nghiệm khách quan, ngân hàng câu hỏi.*

### 1. Đặt vấn đề

Nhằm thực hiện Nghị quyết Hội nghị lần thứ 8, Ban Chấp hành Trung ương khóa XI (Nghị quyết số 29-NQ/TW) với nội dung “Đổi mới căn bản, toàn diện giáo dục và đào tạo, đáp ứng yêu cầu công nghiệp hóa - hiện đại hóa trong điều kiện kinh tế thị trường định hướng xã hội chủ nghĩa và hội nhập quốc tế”, một trong những nội dung quan trọng để nâng cao chất lượng giáo dục chính là việc đổi mới phương pháp dạy học, trong đó có việc đổi mới về phương pháp kiểm tra, đánh giá kết quả học tập của người học đáp ứng yêu cầu đổi mới căn bản, toàn diện giáo dục và đào tạo. Có thể nói việc kiểm tra, đánh giá là hoạt động không thể thiếu của quá trình dạy và học. Kiểm tra, đánh giá kết quả học tập của người học là một vấn đề hết sức quan trọng, bởi vì nó là khâu cuối cùng không những đánh giá độ tin cậy kết quả học tập của quá trình dạy và học mà còn có tác dụng điều tiết trở lại hết sức mạnh mẽ đối với quá trình đào tạo. Thông qua kiểm tra, đánh giá trình độ nhận thức, kỹ năng, kỹ xảo của người học sẽ phát hiện những sai sót, những lỗ hổng về kiến thức để từ đó giúp người dạy và người học tự điều chỉnh hoạt động dạy và hoạt động học. Hướng tới yêu cầu kiểm tra, đánh giá một cách công bằng, khách quan kết quả học tập của người học, để đổi mới phương pháp kiểm tra đánh giá người ta thường sử dụng hình thức thi trắc nghiệm khách quan (TNKQ). Trên thực tế, từ khi tiến hành soạn thảo câu hỏi trắc nghiệm cho đến khi sử dụng được câu hỏi trắc nghiệm vào đề thi thì các câu hỏi trắc nghiệm phải được đánh giá ở nhiều mức độ khác nhau và mỗi đề thi trắc

nghiệm phải bao quát được kiến thức của môn học cần kiểm tra.

Hiện nay, việc sử dụng ngân hàng câu hỏi và thi TNKQ đang được các trường khuyến khích, tuy nhiên đa số các câu hỏi TNKQ do người dạy tự biên soạn có thể chưa theo đúng quy trình, đặc biệt các câu hỏi trước và sau khi sử dụng thường không được phân tích, đánh giá nên phần lớn các đề thi hiện nay là chưa tốt và chất lượng chưa cao. Kết quả nghiên cứu của bài viết này sẽ là tài liệu rất cần thiết góp phần vào quá trình cải thiện chất lượng và hiệu quả của việc biên soạn và thiết kế đề thi TNKQ, đồng thời góp phần nâng cao kỹ năng của giáo viên trong việc thiết kế câu hỏi TNKQ dùng để xây dựng ngân hàng câu hỏi TNKQ. Tiếp theo là phần giới thiệu sơ lược về bảng GSP, phương pháp ROC, và lý thuyết ứng đáp câu hỏi. Năm 2010, Nagai đã đề xuất bảng GSP dựa trên sự kết hợp giữa phân tích quan hệ xám với bảng S-P. Bảng GSP cung cấp thông tin về hệ số chú ý của học sinh và hệ số chú ý của câu hỏi, nó còn tính toán các số liệu rời rạc và định lượng các nhân tố thông qua sắp xếp trình tự để giải quyết các mối liên hệ phức tạp giữa các nhân tố. Ngoài ra, nó còn sử dụng không chỉ để phân tích, chẩn đoán và đánh giá trong học tập, mà còn góp phần nâng cao hiệu quả trong giảng dạy [8]. Trong những năm gần đây, bảng GSP đã được sử dụng nhiều trong lĩnh vực giáo dục [6], [8]. Phương pháp sử dụng đường cong ROC (Receiver Operating Characteristic) có nguồn gốc từ lĩnh vực quân sự, nó được ứng dụng trong việc phát hiện tàu của địch trên màn hình radar trong thế chiến thứ 2. Phương pháp ROC đã được ứng dụng chẩn đoán

<sup>(\*)</sup> Trường Cao đẳng Sư phạm Kiên Giang.

và tiên lượng trong y học rất thành công [5]. Nó cũng được sử dụng trong lĩnh vực giáo dục để phân tích, chẩn đoán và đánh giá trong quá trình dạy học [9]. Lý 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. Lý thuyết ứng đáp câu hỏi xây dựng các mô hình toán để xử lý dữ liệu dựa trên việc nghiên cứu mọi cặp tương tác giữa “Học sinh (HS) - Câu hỏi” khi triển khai một TNKQ. Mỗi HS đứng trước một câu hỏi sẽ ứng đáp như thế nào, điều đó phụ thuộc vào năng lực tiềm ẩn của HS và các đặc trưng của câu hỏi [10]. Hiện nay các lý thuyết về bảng GSP, phương pháp ROC và lý thuyết ứng đáp câu hỏi chưa được sử dụng phổ biến ở Việt Nam, đặc biệt là dùng để phân tích, chẩn đoán và đánh giá trong giáo dục.

Năm 2015, phân tích và lựa chọn câu hỏi TNKQ dựa trên bảng S-P, phân tích quan hệ xám và đường cong ROC đã được đề xuất với cỡ mẫu nhỏ [3]. Tuy nhiên, cỡ mẫu sẽ có ảnh hưởng lớn đến độ chính xác của ước lượng thống kê. Vì vậy, trong bài viết này, người nghiên cứu sử dụng kết hợp bảng GSP và phương pháp ROC để phân tích và lựa chọn câu hỏi TNKQ với cỡ mẫu lớn và so sánh với kết quả của lý thuyết ứng đáp câu hỏi. Lý thuyết ứng đáp câu hỏi là lý thuyết đã được nhiều nước trên thế giới sử dụng phổ biến để phân tích và lựa chọn câu hỏi TNKQ [10]. Hơn nữa, người nghiên cứu còn sử dụng phần mềm MATLAB để hoàn thiện hộp công cụ MATLAB cho phân tích và lựa chọn câu hỏi TNKQ dựa trên bảng GSP và phương pháp ROC. Hộp công cụ MATLAB giúp cho quá trình tính toán dễ dàng, nhanh chóng, chính xác, hiển thị kết quả và hình ảnh trên giao diện đồ họa người dùng một cách trực quan sinh động.

**2. Cơ sở lý thuyết và phương pháp nghiên cứu**

**2.1. Bảng GSP**

Bảng GSP đã được Masatake Nagai đề xuất trong năm 2010 dựa trên sự kết hợp giữa phân tích quan hệ xám và bảng S-P. Bảng S-P được đề xuất bởi Takahiro Sato vào năm 1969. Nó thường dùng để sắp xếp, phân tích và phân loại kết quả học tập của HS và câu hỏi trắc nghiệm dựa trên hệ số chú ý của HS (CS) và hệ số chú ý của câu hỏi (CP) [8].

Trong bảng GSP ở bảng 1,  $Y$  là ma trận có  $m$  hàng và  $n$  cột, trong đó  $y_{ij} = 1$  nếu HS trả lời đúng câu hỏi và  $y_{ij} = 0$  nếu HS trả lời sai câu hỏi. Số HS là  $S_i, i=1,2,\dots,m$ ; số câu hỏi là  $P_j, j=1,2,\dots,n$ .

Hệ số chú ý của HS (CS) được tính bằng công thức sau:

$$CS_i = 1 - \frac{\sum_{j=1}^n (y_{ij})(y_{\bullet j}) - (y_{i\bullet})(\bar{y})}{\sum_{j=1}^n (y_{\bullet j}) - (y_{i\bullet})(\bar{y})}, \text{ trong đó}$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_{\bullet j} \text{ và } N = y_{i\bullet} = \sum_{j=1}^n y_{ij} \quad (1)$$

Hệ số chú ý của câu hỏi (CP) được tính như sau:

$$CP_j = 1 - \frac{\sum_{i=1}^m (y_{ij})(y_{i\bullet}) - (y_{\bullet j})(\bar{y}')}{\sum_{i=1}^m (y_{i\bullet}) - (y_{\bullet j})(\bar{y}')}, \text{ trong đó}$$

$$\bar{y}' = \frac{1}{m} \sum_{i=1}^m y_{i\bullet} \text{ và } M = y_{\bullet j} = \sum_{i=1}^m y_{ij} \quad (2)$$

**Bảng 1. Bảng GSP**

| HS \ Câu hỏi               | Số câu hỏi $P_j, j=1,2,\dots,n$ | Tổng số câu hỏi trả lời đúng | CS     | GS     |
|----------------------------|---------------------------------|------------------------------|--------|--------|
| Số HS $S_i, i=1,2,\dots,m$ | $Y = [y_{ij}]_{m \times n}$     | Cao<br>↓<br>Thấp             | $CS_i$ | $GS_i$ |
| Tổng số HS trả lời đúng    | Nhiều ↔ Ít                      | —                            | —      | —      |
| CP                         | $CP_j$                          | —                            | —      | —      |
| GP                         | $GP_j$                          | —                            | —      | —      |

Trong nghiên cứu này, phân tích quan hệ xám đã được sử dụng dựa theo giá trị lớn nhất (Lager-the-Better) để làm vector tham khảo  $y_0$  [2]. Dựa trên dữ liệu thô từ bảng S-P để thiết lập vector  $y_0$ , vector  $y_0$  là giá trị lớn nhất ở mỗi cột và  $y_i$  là số liệu từng hàng dựa trên dữ liệu thô để so sánh với  $y_0$ .

$$y_0 = (y_0(1), y_0(2), \dots, y_0(k), \dots, y_0(m)) \quad (3)$$

$$\begin{aligned}
 y_1 &= (y_1(1), y_1(2), \dots, y_1(k), \dots, y_1(m)) \\
 y_2 &= (y_2(1), y_2(2), \dots, y_2(k), \dots, y_2(m)) \\
 &\vdots \\
 y_i &= (y_i(1), y_i(2), \dots, y_i(k), \dots, y_i(m)) \quad (4) \\
 &\vdots \\
 y_n &= (y_n(1), y_n(2), \dots, y_n(k), \dots, y_n(m)) \\
 i &= 1, 2, \dots, n
 \end{aligned}$$

Sau khi đã thiết lập được số liệu phân tích thì tiến hành tính toán mức độ quan hệ xám. Công thức tính mức độ quan hệ xám đã được dựa trên lý luận cơ bản về khoảng cách Minkowski. Mức độ quan hệ xám của HS và câu hỏi lần lượt được kí hiệu là  $GS_i$  và  $GP_j$  tương ứng với giá trị  $\Gamma$  của mỗi HS và mỗi câu hỏi. Giá trị  $\Gamma$  về cơ bản được tính như sau [2]:

$$\gamma_{oi} = \gamma(y_0(k), y_i(k)) = \frac{\bar{\Delta}_{\max} - \bar{\Delta}_{oi}}{\bar{\Delta}_{\max} - \bar{\Delta}_{\min}}, i = 1, 2, \dots, n \quad (5)$$

trong đó,  $\bar{\Delta}_{oi}$  là tổng khoảng cách sai số tuyệt đối giữa  $y_i$  với  $y_0$ .

$$\bar{\Delta}_{oi} = \|y_0 - y_i\|_{\rho} = \left( \sum_{j=1}^n (y_0(j) - y_i(j))^{\rho} \right)^{\frac{1}{\rho}} \quad (6)$$

$\bar{\Delta}_{\max}$  và  $\bar{\Delta}_{\min}$  tương ứng là giá trị lớn nhất và giá trị nhỏ nhất của  $\bar{\Delta}_{oi}$ , trong bài viết này người nghiên cứu đã sử dụng  $\rho = 2$  để tính giá trị  $\Gamma$ .

### 2.2. Phương pháp ROC

Phương pháp sử dụng đường cong ROC dùng để đánh giá các kết quả của một dự đoán và ứng dụng đầu tiên của nó là cho việc nghiên cứu các hệ thống nhận diện trong việc phát hiện các tín hiệu radio khi có sự hiện diện của nhiễu vào thập niên 1940 [5]. Trong nghiên cứu này, để xây dựng đường cong ROC các nhà nghiên cứu cần phải tính toán độ nhạy và độ đặc hiệu của từng câu hỏi dựa trên giá trị thực tế và giá trị dự báo để xác định các trạng thái dương tính và âm tính.

**Bảng 2. Bảng 2x2 của phương pháp ROC**

| Giá trị dự báo  | Giá trị thực tế     |                  |
|-----------------|---------------------|------------------|
|                 | Dương tính thật (a) | Âm tính thật (b) |
| Âm tính giả (c) | Dương tính giả (d)  |                  |

Cách xác định trạng thái dương tính và âm tính của câu hỏi như sau: Dựa trên kết quả trả lời đúng sai của HS và căn cứ vào tổng số HS trả lời đúng câu hỏi để xác định trạng thái dương tính (ký hiệu là 1) và âm tính (ký hiệu là 0) của giá trị dự báo. Sau đó căn cứ vào giá trị thực tế để tính các trạng thái  $a, b, c$  và  $d$  của câu hỏi [3].

$$\text{Độ nhạy (Se)} = \frac{a}{a+b} \quad (7)$$

$$\text{Độ đặc hiệu (Sp)} = \frac{d}{c+d} \quad (8)$$

Diện tích bên dưới đường cong

$$(AUC) = \frac{Se(1-Sp)}{2} + \frac{(Se+1)Sp}{2} \quad (9)$$

Đường cong ROC có trục tung là tỉ lệ dương tính thật (độ nhạy) và trục hoành là tỉ lệ dương tính giả (1 trừ cho độ đặc hiệu). Cả hai tỉ lệ này sử dụng xác suất để tính và chúng có giá trị dao động từ 0 đến 1. Theo nhiều nghiên cứu diện tích bên dưới đường cong ROC (AUC) được xem là phân biệt tốt giữa hai trạng thái dương tính và âm tính khi  $AUC \geq 0,7$  [5], [9].

### 2.3. Lý thuyết ứng đáp câu hỏi

Phần mềm BILOG-MG dựa trên lý thuyết ứng đáp câu hỏi đã được rất nhiều nhà nghiên cứu trên thế giới sử dụng để phân tích câu hỏi TNKQ [7]. Hiện nay có 3 mô hình toán phổ biến nhất trong lý thuyết ứng đáp câu hỏi: mô hình 1 tham số chỉ xét đến độ khó của câu hỏi, mô hình 2 tham số có xét đến độ phân biệt của câu hỏi, và mô hình 3 tham số xét thêm mức độ đoán mò của HS khi trả lời câu hỏi. Trong nghiên cứu này, mô hình 3 tham số được sử dụng để phân tích các câu hỏi TNKQ dựa trên phần mềm BILOG-MG. Công thức toán học của mô hình lý thuyết ứng đáp câu hỏi cho 3 tham số như sau:

$$P(\theta) = c + \frac{1-c}{1+e^{-a(\theta-b)}} \quad (10)$$

Trong đó,  $\theta$  là mức độ năng lực,  $a$  là tham số về độ phân biệt,  $b$  là tham số về độ khó và  $c$  là tham số về độ đoán mò. Các tham số này được lựa chọn trong các khoảng sau:  $0,5 \leq a < 2,0$ ;  $-3 \leq b \leq 3$ ; and  $0 \leq c \leq 0,35$ . Các giá trị trong các khoảng của ba tham số trên đã thường được các nhà nghiên cứu sử dụng để phân tích và lựa chọn câu hỏi TNKQ [1].

**2.4. Thiết kế hộp công cụ MATLAB**

Để thuận tiện cho việc tính toán nhanh chóng và chính xác các phép tính phức tạp nhiều nhà nghiên cứu đã sử dụng phần mềm MATLAB để thiết kế một hộp công cụ MATLAB [3], [4], [6]. Trong bài viết này, người nghiên cứu hoàn thiện hộp công cụ MATLAB để phân tích, lựa chọn câu hỏi TNKQ dựa trên bảng GSP và phương pháp ROC, chương trình xử lý dữ liệu của hộp công cụ MATLAB trong nghiên cứu này được tóm tắt gồm có 6 bước (hình 1):

*Bước 1.* Nhập dữ liệu. Dữ liệu là ma trận  $Y$  được nhập vào dưới dạng tập tin \*.csv hoặc \*.xls hoặc \*.xlsx.

*Bước 2.* Kiểm định độ tin cậy của dữ liệu (hệ số Cronbach's Alpha).

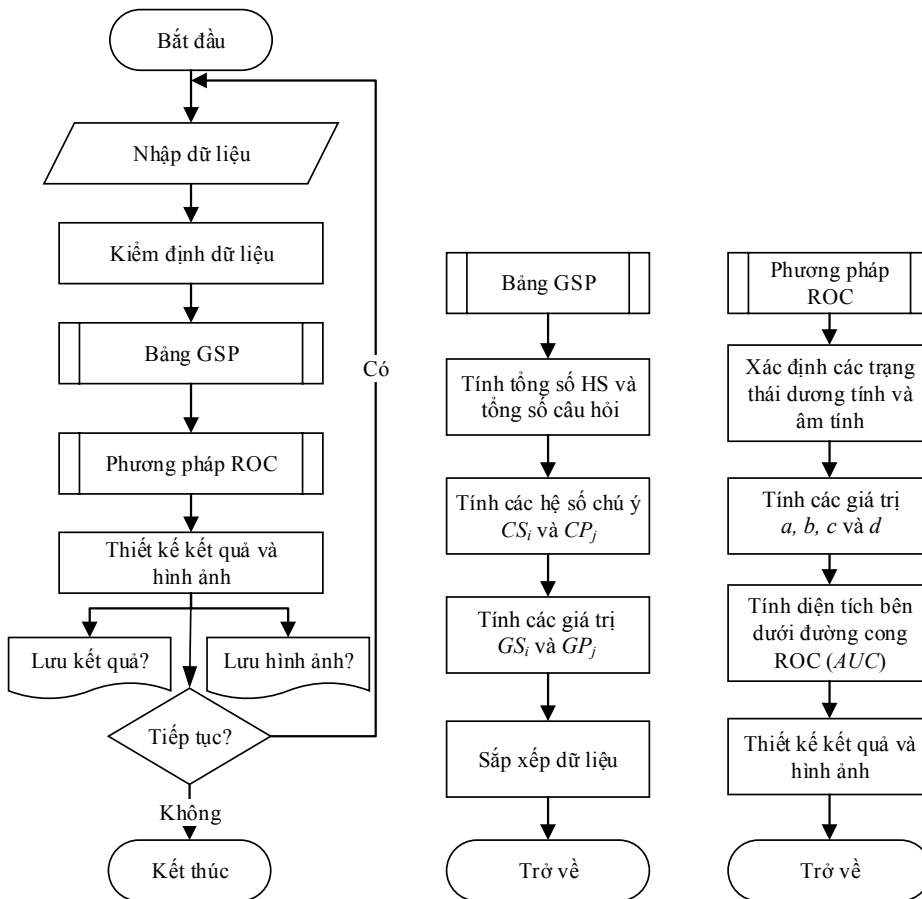
*Bước 3.* Tính tổng số câu hỏi trả lời đúng cho mỗi HS và tổng số HS trả lời đúng ở mỗi câu hỏi; tính hệ số chú ý  $CS$  và  $CP$  của HS và câu hỏi, tiếp theo là sắp xếp theo giá trị  $CS$  và  $CP$  từ nhỏ đến lớn. Thiết lập vector  $y_0$ ; tiếp theo tính

tổng khoảng cách sai số tuyệt đối của mỗi HS và mỗi câu hỏi; tính giá trị  $\Gamma$  ( $GS$  và  $GP$ ) của mỗi HS và mỗi câu hỏi; sau đó sắp xếp theo giá trị  $\Gamma$  từ lớn đến nhỏ đối với  $GS$  và  $GP$ , thiết kế kết quả và hình ảnh cho các giá trị  $\Gamma$ .

*Bước 4.* Xác định các trạng thái dương tính và âm tính; tính các giá trị  $a, b, c$  và  $d$ ; tiếp theo, tính diện tích bên dưới đường cong ROC của mỗi câu hỏi; sau đó thiết kế kết quả và hình ảnh đường cong ROC.

*Bước 5.* Thiết kế hiển thị các kết quả và hình ảnh để hiển thị trên giao diện đồ họa người dùng. Người sử dụng có thể lưu lại kết quả dưới dạng tập tin \*.csv hoặc \*.xls hoặc \*.xlsx và hình ảnh dưới dạng tập tin \*.JPG.

*Bước 6.* Tiếp tục hoặc thoát khỏi chương trình. Nếu người sử dụng nhập dữ liệu mới thì chương trình sẽ tiếp tục và trở về bước 1, hoặc thoát khỏi chương trình thì chương trình sẽ đóng lại.



**Hình 1.** Lưu đồ phân tích câu hỏi TNKQ dựa trên bảng GSP và phương pháp ROC

### 3. Kết quả nghiên cứu và thảo luận

#### 3.1. Dữ liệu thực nghiệm

Dữ liệu trong nghiên cứu này là kết quả trả lời 25 câu hỏi trắc nghiệm Toán học của 813 HS lớp 7. Dữ liệu và phần mềm BILOG-MG sử dụng trong nghiên cứu này được lấy từ Viện Thông tin và Đo lường Giáo dục, Trường Đại học Sư phạm Đài Trung, Đài Loan (Graduate Institute of Educational Information and Measurement, National Taichung University of Education, Taiwan). Trước khi tiến hành phân tích lựa chọn câu hỏi trắc nghiệm, người nghiên cứu đã kiểm tra độ tin cậy của dữ liệu thông qua hệ số Cronbach's Alpha. Hệ số Cronbach's Alpha của dữ liệu là

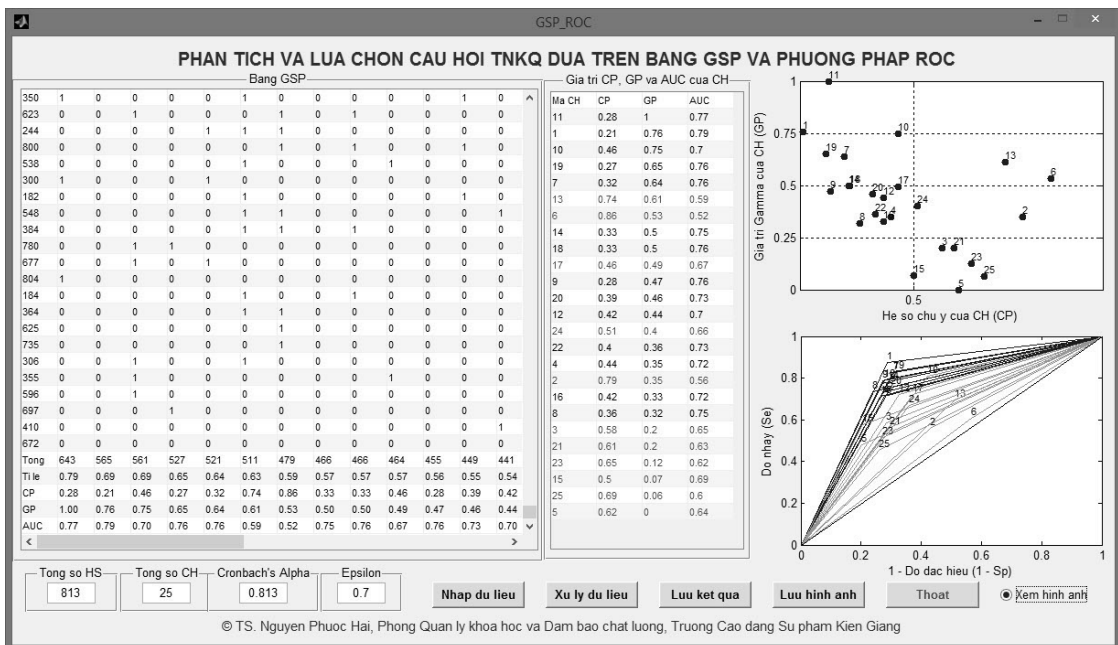
0,813, hệ số này cho thấy dữ liệu có độ tin cậy cao, phù hợp để thực hiện nghiên cứu.

#### 3.2. Kết quả nghiên cứu

Trong nghiên cứu này, mô hình 3 tham số dựa trên phần mềm BILOG-MG đã được sử dụng để phân tích 25 câu hỏi TNKQ. Kết quả phân tích và lựa chọn câu hỏi dựa trên lý thuyết ứng đáp câu hỏi cho thấy câu hỏi 13, câu hỏi 6 và câu hỏi 2 không được chấp nhận theo điều kiện của lý thuyết ứng đáp câu hỏi (hình 2 và bảng 3). Bởi vì, câu hỏi 13 có tham số độ phân biệt là 0,28 ( $a$ ); câu hỏi 6 có tham số độ đoán mò là 0,50 ( $c$ ); và câu hỏi 2 có tham số độ phân biệt ( $a$ ) là 0,39 và tham số độ đoán mò là 0,38 ( $c$ ) không đáp ứng điều kiện của lý thuyết ứng đáp câu hỏi [1].

| Item | Type | Difficulty | Discrimination | Guessing | Other   | Other    | Other   | Other   | Other   |         |         |         |         |      |
|------|------|------------|----------------|----------|---------|----------|---------|---------|---------|---------|---------|---------|---------|------|
| 1    | TEST | 0.82992    | 0.09996        | 1.67567  | 0.17401 | -0.49528 | 0.07333 | 0.85871 | 0.08917 | 0.09575 | 0.03833 | 0.00000 | 0.00000 | 110  |
| 2    | TEST | -0.99655   | 0.32066        | 0.38862  | 0.11893 | 2.56431  | 0.64265 | 0.36223 | 0.11085 | 0.38220 | 0.05160 | 0.00000 | 0.00000 | 210  |
| 3    | TEST | -0.90483   | 0.23639        | 0.78929  | 0.16823 | 1.14638  | 0.15179 | 0.61956 | 0.13205 | 0.21911 | 0.04228 | 0.00000 | 0.00000 | 310  |
| 4    | TEST | -0.32720   | 0.14288        | 0.85841  | 0.13030 | 0.38117  | 0.12782 | 0.65135 | 0.09887 | 0.15233 | 0.04787 | 0.00000 | 0.00000 | 410  |
| 5    | TEST | -2.53565   | 0.66832        | 1.69375  | 0.48939 | 1.49705  | 0.11400 | 0.86112 | 0.24881 | 0.20193 | 0.02034 | 0.00000 | 0.00000 | 510  |
| 6    | TEST | -1.48556   | 0.45560        | 0.50647  | 0.17452 | 2.93316  | 0.81350 | 0.45183 | 0.15569 | 0.50000 | 0.03414 | 0.00000 | 0.00000 | 610  |
| 7    | TEST | 0.25630    | 0.11686        | 1.12164  | 0.14364 | -0.22851 | 0.12170 | 0.74642 | 0.09559 | 0.16761 | 0.05512 | 0.00000 | 0.00000 | 710  |
| 8    | TEST | -0.56369   | 0.16708        | 1.34111  | 0.20202 | 0.42032  | 0.08130 | 0.80167 | 0.12076 | 0.16776 | 0.03449 | 0.00000 | 0.00000 | 810  |
| 9    | TEST | -0.14993   | 0.14407        | 1.56302  | 0.24830 | 0.09592  | 0.08168 | 0.84235 | 0.13382 | 0.16661 | 0.04076 | 0.00000 | 0.00000 | 910  |
| 10   | TEST | 0.31232    | 0.13243        | 0.72116  | 0.10576 | -0.49308 | 0.22321 | 0.58432 | 0.08578 | 0.22188 | 0.07950 | 0.00000 | 0.00000 | 1010 |
| 11   | TEST | 1.06050    | 0.12087        | 1.24490  | 0.14574 | -0.85229 | 0.13291 | 0.77947 | 0.09129 | 0.18362 | 0.07067 | 0.00000 | 0.00000 | 1110 |
| 12   | TEST | -0.35851   | 0.17655        | 1.13380  | 0.19899 | 0.31620  | 0.11532 | 0.74997 | 0.13162 | 0.22542 | 0.04757 | 0.00000 | 0.00000 | 1210 |
| 13   | TEST | -0.14767   | 0.20018        | 0.27617  | 0.06940 | 0.53471  | 0.67048 | 0.26621 | 0.06111 | 0.33252 | 0.09194 | 0.00000 | 0.00000 | 1310 |
| 14   | TEST | -0.09105   | 0.14268        | 1.26109  | 0.19819 | 0.07220  | 0.10476 | 0.78355 | 0.12314 | 0.17921 | 0.04813 | 0.00000 | 0.00000 | 1410 |
| 15   | TEST | -2.17103   | 0.58612        | 1.94729  | 0.54312 | 1.11490  | 0.07529 | 0.88956 | 0.24811 | 0.19707 | 0.02187 | 0.00000 | 0.00000 | 1510 |
| 16   | TEST | -0.25001   | 0.10617        | 0.82510  | 0.09878 | 0.30301  | 0.10805 | 0.63643 | 0.07619 | 0.10141 | 0.03827 | 0.00000 | 0.00000 | 1610 |
| 17   | TEST | -0.60302   | 0.24671        | 1.23246  | 0.27494 | 0.48928  | 0.11867 | 0.77654 | 0.17323 | 0.33041 | 0.04500 | 0.00000 | 0.00000 | 1710 |
| 18   | TEST | -0.05213   | 0.13078        | 1.11753  | 0.15816 | 0.04665  | 0.11212 | 0.74521 | 0.10547 | 0.16718 | 0.04880 | 0.00000 | 0.00000 | 1810 |
| 19   | TEST | 0.27894    | 0.12818        | 1.65766  | 0.27110 | -0.16827 | 0.09256 | 0.85626 | 0.14003 | 0.19841 | 0.04920 | 0.00000 | 0.00000 | 1910 |
| 20   | TEST | -0.15324   | 0.14442        | 0.99533  | 0.16133 | 0.15396  | 0.12725 | 0.70545 | 0.11434 | 0.17462 | 0.05332 | 0.00000 | 0.00000 | 2010 |
| 21   | TEST | -1.29761   | 0.32111        | 1.01501  | 0.23907 | 1.27842  | 0.13602 | 0.71235 | 0.16778 | 0.27113 | 0.03316 | 0.00000 | 0.00000 | 2110 |
| 22   | TEST | -0.74821   | 0.22637        | 1.53656  | 0.28293 | 0.46694  | 0.08140 | 0.83813 | 0.15433 | 0.23591 | 0.03458 | 0.00000 | 0.00000 | 2210 |
| 23   | TEST | -1.64990   | 0.40820        | 1.07673  | 0.28715 | 1.53233  | 0.15765 | 0.73273 | 0.19541 | 0.26113 | 0.02886 | 0.00000 | 0.00000 | 2310 |
| 24   | TEST | -0.98603   | 0.31163        | 1.36854  | 0.33043 | 0.72050  | 0.10018 | 0.80742 | 0.19495 | 0.32750 | 0.03666 | 0.00000 | 0.00000 | 2410 |
| 25   | TEST | -2.33613   | 0.64546        | 1.30524  | 0.41134 | 1.78981  | 0.18341 | 0.79381 | 0.25017 | 0.26243 | 0.02313 | 0.00000 | 0.00000 | 2510 |

Hình 2. Kết quả phân tích 25 câu hỏi dựa trên phần mềm BILOG-MG



Hình 3. Giao diện đồ họa người dùng cho phân tích và lựa chọn 25 câu hỏi TNKQ

Trên giao diện đồ họa, người dùng của hộp công cụ MATLAB dùng để phân tích và lựa chọn câu hỏi TNKQ dựa trên bảng GSP và phương pháp ROC (GSP-ROC), có thể thấy kết quả các giá trị  $CP$ ,  $Gamma$  và  $AUC$  của các câu hỏi và hình ảnh về giá trị  $Gamma$ , các đường cong ROC của câu hỏi trong quá trình phân tích và lựa chọn câu hỏi TNKQ. Theo kết quả phân tích và lựa chọn câu hỏi TNKQ dựa trên bảng GSP và phương pháp ROC (hình 3 và bảng 3), nếu chọn ngưỡng giá trị  $Epsilon$  bằng 0,6 thì kết quả giống với kết quả của phương pháp lựa chọn câu hỏi dựa theo lý thuyết ứng đáp câu hỏi, nghĩa là có 3 câu hỏi không được chấp nhận (câu hỏi 13, câu hỏi 6 và câu hỏi 2). Tuy nhiên, nếu chọn ngưỡng giá trị  $Epsilon$  bằng 0,7 thì sẽ có 11 câu hỏi không được chấp nhận trong nghiên cứu này.

**Bảng 3. Kết quả phân tích 25 câu hỏi dựa trên BILOG-MG và GSP-ROC**

| Câu hỏi | BILOG-MG    |       |             | GSP-ROC |         |             | Kết quả |
|---------|-------------|-------|-------------|---------|---------|-------------|---------|
|         | $a$         | $b$   | $c$         | $CP$    | $Gamma$ | $AUC$       |         |
| 11      | 1,24        | -0,85 | 0,18        | 0,28    | 1,00    | 0,77        | Đạt     |
| 1       | 1,68        | -0,50 | 0,10        | 0,21    | 0,76    | 0,79        | Đạt     |
| 10      | 0,72        | -0,43 | 0,22        | 0,46    | 0,75    | 0,70        | Đạt     |
| 19      | 1,66        | -0,17 | 0,20        | 0,27    | 0,65    | 0,76        | Đạt     |
| 7       | 1,12        | -0,23 | 0,17        | 0,32    | 0,64    | 0,76        | Đạt     |
| 13      | <b>0,28</b> | 0,53  | 0,33        | 0,74    | 0,61    | <b>0,59</b> | Không   |
| 6       | 0,51        | 2,93  | <b>0,50</b> | 0,86    | 0,53    | <b>0,52</b> | Không   |
| 14      | 1,26        | 0,07  | 0,18        | 0,33    | 0,50    | 0,75        | Đạt     |
| 18      | 1,12        | 0,05  | 0,17        | 0,33    | 0,50    | 0,76        | Đạt     |
| 17      | 1,23        | 0,49  | 0,33        | 0,46    | 0,49    | <b>0,67</b> | Không   |
| 9       | 1,56        | 0,10  | 0,17        | 0,28    | 0,47    | 0,76        | Đạt     |
| 20      | 1,00        | 0,15  | 0,17        | 0,39    | 0,46    | 0,73        | Đạt     |
| 12      | 1,13        | 0,32  | 0,23        | 0,42    | 0,44    | 0,70        | Đạt     |
| 24      | 1,37        | 0,72  | 0,33        | 0,51    | 0,40    | <b>0,66</b> | Không   |
| 22      | 1,54        | 0,49  | 0,24        | 0,40    | 0,36    | 0,73        | Đạt     |
| 4       | 0,86        | 0,38  | 0,15        | 0,44    | 0,35    | 0,72        | Đạt     |
| 2       | <b>0,39</b> | 2,56  | <b>0,38</b> | 0,79    | 0,35    | <b>0,56</b> | Không   |
| 16      | 0,83        | 0,30  | 0,10        | 0,42    | 0,33    | 0,72        | Đạt     |
| 8       | 1,34        | 0,42  | 0,17        | 0,36    | 0,32    | 0,75        | Đạt     |
| 3       | 0,79        | 1,15  | 0,22        | 0,58    | 0,20    | <b>0,65</b> | Không   |
| 21      | 1,02        | 1,28  | 0,27        | 0,61    | 0,20    | <b>0,63</b> | Không   |
| 23      | 1,08        | 1,53  | 0,26        | 0,65    | 0,12    | <b>0,62</b> | Không   |
| 15      | 1,95        | 1,11  | 0,20        | 0,50    | 0,07    | <b>0,69</b> | Không   |
| 25      | 1,31        | 1,79  | 0,26        | 0,69    | 0,06    | <b>0,60</b> | Không   |
| 5       | 1,69        | 1,50  | 0,20        | 0,62    | 0,00    | <b>0,64</b> | Không   |

“ $a$ ” là tham số về độ phân biệt, “ $b$ ” là tham số về độ khó, “ $c$ ” là tham số về độ đoán mò

### 3.3. Thảo luận

Trong nghiên cứu này, phương pháp đề xuất đã được so sánh với lý thuyết ứng đáp câu hỏi. Lý thuyết ứng đáp câu hỏi 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. Lý thuyết này đã được nhiều nước trên thế giới sử dụng phổ biến và mang lại kết quả rất tích cực [10]. Vì vậy, kết quả phân tích và lựa chọn câu hỏi TNKQ của phương pháp đề xuất là phù hợp và đáng tin cậy để áp dụng cho việc xây dựng ngân hàng câu hỏi TNKQ. Dựa vào bảng kết quả ở bảng 3 có thể thấy rằng hệ số  $CP$  càng cao thì giá trị  $AUC$  càng thấp và ngược lại. Khi hệ số  $CP$  càng cao và giá trị  $AUC$  càng thấp thì điều này cho thấy câu hỏi tương ứng có độ phân biệt thấp. Bởi vì, khi đó HS có năng lực thấp lại trả lời đúng các câu hỏi khó. Do đó, phương pháp nghiên cứu này giúp cho việc xác định và loại bỏ các câu hỏi có hệ số  $CP$  cao và giá trị  $AUC$  thấp.

Hộp công cụ MATLAB trong nghiên cứu này cũng cho thấy là rất hữu dụng và tiện ích, giúp xác định vị trí của các câu hỏi dựa theo hệ số chú ý  $CP$  và giá trị  $GP$  (hình 3), từ đó giúp cho việc loại bỏ các câu hỏi có hệ số chú ý  $CP$  lớn hơn 0,5. Dựa trên phương pháp này cũng có thể tính được hệ số chú ý  $CS$ , giá trị  $Gamma$  và giá trị  $AUC$  của từng HS. Trên cơ sở đó có thể xác định được năng lực học tập của mỗi HS. Ngoài ra, nó cũng cung cấp cho giáo viên những thông tin cần thiết nhằm xác định đúng hơn về nhận thức của HS trong học tập, từ đó đề xuất kịp thời các biện pháp điều chỉnh hoạt động dạy học, thực hiện mục đích dạy học đáp ứng yêu cầu đổi mới căn bản, toàn diện giáo dục và đào tạo.

### 4. Kết luận

Kết quả nghiên cứu cho thấy có thể sử dụng bảng GSP và phương pháp ROC để phân tích và lựa chọn câu hỏi TNKQ đối với trường hợp cỡ mẫu lớn. Bài viết này sẽ rất hữu ích cho các nhà quản lý giáo dục, giáo viên và tất cả những ai quan tâm đến việc tìm kiếm một phương pháp để nâng cao hiệu quả của việc kiểm tra, đánh giá HS thông qua câu hỏi TNKQ.

Nghiên cứu này đã bổ sung, hoàn thiện hộp công cụ MATLAB cho phân tích và lựa chọn câu hỏi TNKQ dựa trên bảng GSP và phương pháp ROC. Trên giao diện đồ họa người dùng của hộp

công cụ MATLAB có thể xác định được vị trí của các câu hỏi dựa theo hệ số chú ý  $CP$  và giá trị  $GP$ , giúp cho việc phân tích và lựa chọn câu hỏi dễ dàng hơn.

Từ những kết quả nghiên cứu cho thấy, đây là phương pháp mà các giáo viên có thể sử

dụng nhằm cải thiện chất lượng biên soạn và thiết kế đề thi, đề kiểm tra TNKQ, đồng thời nghiên cứu này cũng cho thấy giáo viên có thể áp dụng phương pháp này xây dựng ngân hàng câu hỏi TNKQ để kiểm tra, đánh giá kết quả học tập của HS.

### Tài liệu tham khảo

- [1]. Baker, F. B. (2001), "The basics of item response theory", <http://ericae.net/irt/baker>.
- [2]. Nguyễn Phước Hải, Dư Thống Nhất (2014), "Đánh giá kết quả xếp hạng và dự báo kết quả học tập của HS dựa trên phân tích quan hệ xám và mô hình xám", *Tạp chí khoa học Trường Đại học Cần Thơ*, (Số 32), tr. 43-50.
- [3]. Nguyễn Phước Hải, Dư Thống Nhất (2015), "Phân tích và lựa chọn câu hỏi TNKQ dựa trên bảng S-P, phân tích quan hệ xám và đường cong ROC", *Tạp chí Khoa học Đại học Sư phạm Thành phố Hồ Chí Minh*, (Số 6 (72)), tr. 163-173.
- [4]. Nguyễn Phước Hải, Sheu, T. W., & Nagai, M. (2015), "Dự báo kết quả học tập của HS dựa trên sự kết hợp phương pháp gần đúng Taylor và các mô hình xám", *Tạp chí Khoa học Đại học Quốc gia Hà Nội: Nghiên cứu Giáo dục*, (Số 31(2)), tr. 70-83.
- [5]. Kumar, R., & Indrayan, A. (2011), "Receiver operating characteristic (ROC) curve for medical researchers", *Indian Pediatrics*, 48(4), 277-287.
- [6]. Nguyen, P. H., Sheu, T. W., Nguyen, P. T., Pham, D. H., & Nagai, M. (2014), "Taylor Approximation Method in Grey System Theory and Its Application to Predict the Number of Teachers and Students for Admission", *International Journal of Innovation and Scientific Research*, 10(2), 353-363.
- [7]. Sheu, T. W., Nguyen, P. T., Tsai, C. P., Pham, D. H., Nguyen, P. H., & Nagai, M. (2014), "Using Grey Student-Problem Chart in the Evaluation of Tests with Large Data Sets", *Education Practice and Innovation*, 1(2), 40-50.
- [8]. Rupp, A. A. (2003), "Item response modeling with BILOG-MG and MULTILOG for Windows", *International Journal of Testing*, 3(4), 365-384.
- [9]. Tavakol, M., & Dennick, R. (2012), "Standard Setting: the application of the Receiver Operating Characteristic method", *International Journal of Medical Education*, 3, 198-200.
- [10]. Lâm Quang Thiệp (2011), *Đo lường trong giáo dục - Lí thuyết và ứng dụng*, NXB Đại học Quốc gia Hà Nội, Hà Nội.

### USING GSP CHART AND ROC METHOD TO ANALYZE AND SELECT MULTIPLE CHOICE ITEMS

#### Summary

This paper aims to propose the analysis - selection method for multiple-choice items based on GSP chart and ROC method with large samples. In this study, the proposed method has been compared with that of the item-response. The research results showed that this method can be used not only to analyze and select multiple-choice items, but also improve the quality and efficiency of designing test items to build a test item bank related.

Keywords: GSP chart, ROC method, item response theory, multiple-choice items, item bank.

Ngày nhận bài: 10/3/2016; Ngày nhận lại: 26/6/2016; Ngày duyệt đăng: 27/9/2016.