

DOI:10.22144/ctu.jsi.2020.090

## SỬ DỤNG PHẦN MỀM THỐNG KÊ TRONG VIỆC SO BẰNG ĐỀ KIỂM TRA TRẮC NGHIỆM KHÁCH QUAN MÔN TOÁN LỚP 12

Bùi Anh Kiệt<sup>1\*</sup> và Tăng Hòa Thái<sup>2</sup>

<sup>1</sup>Khoa Sư phạm, Trường Đại học Cần Thơ

<sup>2</sup>Sở Giáo dục và Khoa học Công nghệ Bạc Liêu

\*Người chịu trách nhiệm về bài viết: Bùi Anh Kiệt (email: bakiet@ctu.edu.vn)

### Thông tin chung:

Ngày nhận bài: 04/03/2020

Ngày nhận bài sửa: 24/03/2020

Ngày duyệt đăng: 29/06/2020

### Title:

Using statistical softwares to equate the mathematical tests of the multiple – choice questions in grade 12

### Từ khóa:

So bằng đề kiểm tra, lý thuyết ứng đáp câu hỏi, điểm thô, điểm năng lực, đề trắc nghiệm khách quan

### Keywords:

Equating the test, Item Response Theory, multiple – choice questions, the raw score, the latent trait

### ABSTRACT

In this article, the theory of equating the test is introduced and a procedure to write, analyze, and equate two mathematical tests of the multiple – choice questions are suggested. Statistical softwares (IATA and Excel) and Item Response Theory (IRT) were used in order to establish a common scale for two tests. These are the base to transfer the raw score as well as the latent trait or the ability score  $\theta$  of testees from one test to another, in order to equate the difficulty between two tests and assess exactly the ability of testees. Using the above theory, this study has experimented two one-period tests for one chapter of grade 12 mathematical curriculum for students in Bac Lieu, Soc Trang, and Kien Giang provinces in 2019.

### TÓM TẮT

Bài viết giới thiệu về lý thuyết so bằng đề kiểm tra; đề xuất một qui trình biên soạn, thẩm định và so bằng hai đề kiểm tra trắc nghiệm khách quan môn Toán. Qui trình này sử dụng các phần mềm thống kê (IATA và Excel) và lý thuyết ứng đáp câu hỏi (IRT) để thiết lập một thang đo chung cho hai đề kiểm tra. Đây là cơ sở để chuyển điểm thô và điểm năng lực  $\theta$  của thí sinh từ đề kiểm tra này sang đề kiểm tra kia, nhằm mục đích so bằng độ khó giữa hai đề kiểm tra và đánh giá đúng năng lực của thí sinh. Ứng dụng lý thuyết nói trên, nghiên cứu này thực nghiệm trên hai đề kiểm tra một tiết cho một chương của chương trình toán lớp 12 ở một số học sinh của ba tỉnh Bạc Liêu, Sóc Trăng và Kiên Giang năm học 2019.

Trích dẫn: Bùi Anh Kiệt và Tăng Hòa Thái, 2020. Sử dụng phần mềm thống kê trong việc so bằng đề kiểm tra trắc nghiệm khách quan môn Toán lớp 12. Tạp chí Khoa học Trường Đại học Cần Thơ. 56(Số chuyên đề: Khoa học tự nhiên)(1): 28-35.

### 1 ĐẶT VẤN ĐỀ

Trong những năm gần đây, hình thức kiểm tra, đánh giá kết quả học tập bằng trắc nghiệm khách quan (TNKQ) được sử dụng ngày càng phổ biến ở nước ta, cụ thể từ năm 2017, đề thi Trung học phổ thông quốc gia (THPTQG) môn Toán theo hình thức TNKQ. Đây là hình thức thi có nhiều ưu điểm,

nhưng hạn chế của nó là không thể đánh giá một cách chính xác khả năng tư duy của học sinh (HS) (Lâm Quang Thiệp, 2011). Bên cạnh đó, trong đề thi THPTQG môn Toán, mỗi HS phải làm một mã đề trong số 24 mã đề, do đó vấn đề đặt ra là “liệu các mã đề đó có độ khó tương đương nhau hay không?

Có đảm bảo công bằng cho các thí sinh (TS) khi làm các đề kiểm tra (ĐKT) đó hay không?”.

Để trả lời cho các câu hỏi này, một quy trình so bằng ĐKT được trình bày dựa trên cơ sở lý thuyết so bằng theo lý thuyết ứng đáp câu hỏi (IRT), sử dụng phần mềm IATA và Excel. So bằng ĐKT sẽ giúp kết luận được về độ khó của hai ĐKT có tương đương nhau hay không, đây là cơ sở để bảo đảm công bằng cho TS.

## 2 SƠ LƯỢC VỀ LÝ THUYẾT ỨNG ĐÁP CÂU HỎI (IRT)

Trong nửa cuối thế kỉ 20, nhiều nhà nghiên cứu về đo lường trong tâm lý và giáo dục đã cố gắng tìm giải pháp nhằm nâng cao độ chính xác và tính khách quan của các phép đo dùng TNKQ. Trong một bài kiểm tra hoặc một bài thi, việc HS trả lời câu hỏi đúng hay sai phụ thuộc vào năng lực của HS đó trong phạm vi được kiểm tra và độ khó của từng câu hỏi cụ thể. Lý thuyết ứng đáp câu hỏi (IRT) là một phương pháp được dùng để tính xác suất một HS trả lời đúng một câu hỏi thông qua mối tương quan giữa năng lực HS và các tham số đặc trưng của câu hỏi. Ở nước ngoài, có nhiều học giả đã nghiên cứu chuyên sâu về IRT như Linda and James (2008) đã trình bày về các lý thuyết đo lường, độ tin cậy, độ giá trị, IRT,... Thomas (2004) nghiên cứu về TNKQ và ứng dụng IRT vào TNKQ. Ayala (2009) cũng đã trình bày một cách tương đối chi tiết về IRT (có sự so sánh với lý thuyết khảo thí cổ điển) và ứng dụng vào phân tích các câu hỏi TNKQ. Ở Việt Nam, Lâm Quang Thiệp (2011), Dương Thiệu Tống (2005),... cũng có nhiều công trình nghiên cứu về ứng dụng IRT trong kiểm tra, đánh giá giáo dục.

Có hai mô hình IRT thường được sử dụng là mô hình một tham số và mô hình hai tham số.

### 2.1 Mô hình IRT một tham số

Theo Ayala (2009), mô hình IRT đơn giản nhất là mô hình một tham số hay còn gọi là mô hình Rasch. Nhà toán học Đan Mạch, Georg Rasch, đã đưa ra một mô hình “ứng đáp câu hỏi” để mô tả mối tương tác nguyên tố giữa một HS với một câu hỏi của ĐTN, và dùng mô hình đó để phân tích các dữ liệu thu được từ ĐTN. Theo mô hình Rasch, xác suất để HS  $i$  trả lời đúng câu hỏi  $j$  là

$$P_j(\theta_i) = \frac{e^{\theta_i - b_j}}{1 + e^{\theta_i - b_j}}, \text{ trong đó } \theta_i \text{ là năng lực của HS}$$

$i$  và  $b_j$  là độ khó của câu hỏi  $j$ . Giá trị của  $\theta$  tỉ lệ thuận với năng lực của HS và giá trị của  $b$  càng lớn khi và chỉ khi câu hỏi càng khó. Một HS muốn có xác suất trả lời đúng một câu hỏi là 0,5 thì HS đó phải có năng lực bằng với độ khó của câu hỏi đó, cụ thể là  $\theta_i = b_j$ . Năng lực của HS thường được giả sử là một phân phối chuẩn với kì vọng là 0 và phương sai là 1. Tập giá trị thường gặp của năng lực  $\theta$  là từ -3 đến 3 hoặc từ -4 đến 4. Các giá trị bên ngoài khoảng này cho thấy câu hỏi có vấn đề.

### 2.2 Mô hình IRT hai tham số

Bên cạnh độ khó, chúng ta có thể đưa thêm một tham số thứ hai liên quan đến độ phân biệt của câu hỏi là  $a_j$  vào mô hình, và ta thường gọi nó là mô hình hai tham số. Theo mô hình đó, xác suất để HS

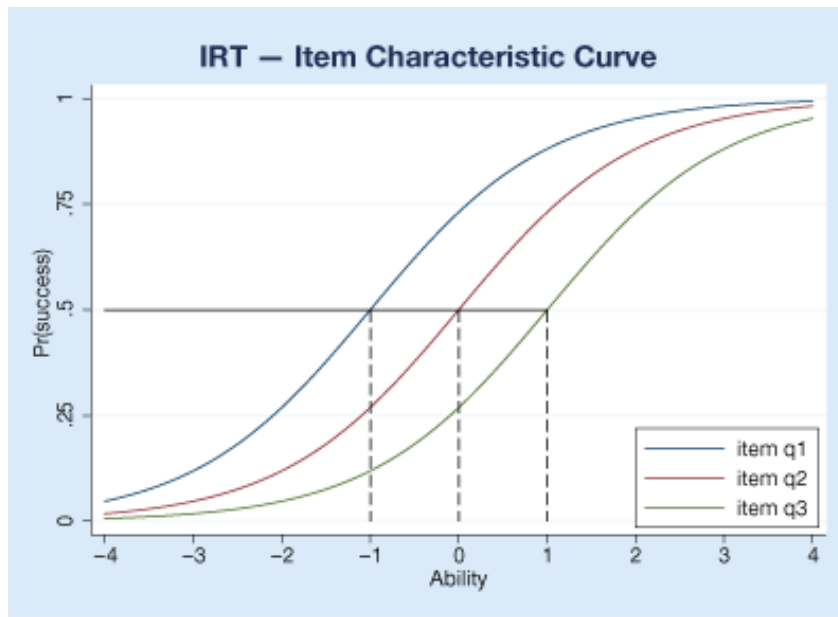
$$i \text{ trả lời đúng câu hỏi } j \text{ là } P_j(\theta_i) = \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}},$$

trong đó  $\theta_i, b_j$  có ý nghĩa như trong mô hình một tham số và  $a_j$  là độ phân biệt của câu hỏi thứ  $j$

(Ayala, 2009). Giá trị của  $a_j$  thường nằm trong khoảng từ 0 đến 2. Khi độ phân biệt có giá trị càng lớn thì một sự khác biệt rất nhỏ về năng lực cũng sẽ gây ra một độ chênh lệch lớn về xác suất trả lời đúng. Dễ dàng nhận thấy rằng nếu  $a_j = 1$  với mọi  $j$  thì ta thu được mô hình Rasch. Tương tự như độ khó, một câu hỏi có độ phân biệt bằng các giá trị biên hoặc vượt khỏi khoảng từ 0 đến 2 cho thấy câu hỏi có thể có vấn đề về nội dung hoặc cách diễn đạt. Đặc biệt, độ phân biệt âm chỉ rằng khả năng trả lời đúng của những HS có năng lực cao lại nhỏ hơn so với những HS có năng lực thấp. Những câu hỏi này cần được xem xét lại để đảm bảo rằng chúng phù hợp với nội dung mà chúng ta đang kiểm tra.

### 2.3 Đường cong đặc trưng câu hỏi

Nếu cho giá trị của  $\theta$  biến đổi liên tục trên trục số (giả sử từ -4 đến 4) thì chúng ta thu được một đồ thị được gọi là **đường cong đặc trưng câu hỏi (Item Characteristic Curve – ICC)** có dạng như sau:



**Hình 1: Đường cong đặc trưng câu hỏi theo mô hình Rasch**

Chúng ta có thể tổng kết các tính chất của một câu hỏi bằng cách nhìn vào ICC. ICC thể hiện xác suất trả lời đúng một câu hỏi ứng với các giá trị khác nhau của  $\theta$  và nó là một đường cong logistic có biên từ 0 đến 1. Độ khó một câu hỏi chính là tọa độ điểm uốn của ICC theo thang năng lực  $\theta$ . Theo Lâm Quang Thiệp (2011), nếu kẻ một đường thẳng song song với trục hoành và cắt trục tung tại vị trí  $P = 0,5$  thì đường thẳng này sẽ cắt các ICC tại các điểm có hoành độ đúng bằng độ khó của câu hỏi tương ứng, vì điểm của đường cong ứng với tung độ  $P = 0,5$  sẽ có hoành độ  $\theta = b$ . Các đường cong càng dịch chuyển về phía bên trái (phải) biểu thị cho câu hỏi càng dễ (khó). Ngoài ra, độ phân biệt mô tả độ dốc của ICC tại điểm có hoành độ  $\theta = b$ , tung độ  $P(\theta) = 0,5$ . Một ICC có độ dốc (hệ số góc) càng lớn cho thấy câu hỏi đó có khả năng phân biệt năng lực HS càng tốt. Một ICC có hệ số góc âm khi và chỉ khi độ phân biệt âm. Các ICC càng gần với đồ thị hàm bậc thang biểu thị cho câu hỏi có độ phân biệt đặc biệt lớn.

### 3 LÝ THUYẾT SO BẰNG ĐỀ KIỂM TRA

#### 3.1 Khái niệm so bằng

Theo Lâm Quang Thiệp (2011), “So bằng là qui trình tìm một hàm nào đó để chuyển điểm của thí sinh thu được từ một đề khảo sát nào đó sang một thang điểm của đề quy ước làm gốc” (trang 180).

Finch (2019) đã trình bày về các lý thuyết so bằng, so bằng về điểm thô và so bằng điểm năng lực

$Q$  theo IRT. Ở Việt Nam, Lâm Quang Thiệp (2011) đã trình bày cơ bản về việc so bằng hai đề trắc nghiệm (ĐTN) tiếng Việt năm 2001 (VIỆT01) và năm 2007 (VIỆT07).

Có hai phương pháp so bằng (equating) ĐTN, đó là so bằng theo CTT và so bằng kết nối - xác lập thang đo theo IRT. Chúng tôi chọn phương pháp so bằng kết nối - xác lập thang đo theo IRT vì hiện nay hầu như các nhà nghiên cứu về đo lường và đánh giá đều sử dụng IRT. Theo IRT, so bằng chủ yếu là về điểm thô và điểm năng lực  $\theta$  theo IRT. Như vậy, so bằng hai ĐKT là so bằng về điểm thô và điểm năng lực  $\theta$  theo IRT.

#### 3.2 Thiết kế kết nối tạo lập thang đo

Theo Lâm Quang Thiệp (2011) có rất nhiều cách thiết kế, kết nối để tạo lập thang đo các tham số câu hỏi như: thiết kế đơn nhóm, thiết kế các nhóm tương đương, thiết kế các ĐTN có các câu hỏi neo (câu hỏi chung của hai đề kiểm tra) và thiết kế có các TS chung.

Trong các thiết kế được mô tả trên, theo Lâm Quang Thiệp (2011): “thiết kế ĐTN có các câu hỏi neo là khả thi nhất” (trang 185).

Trong các thiết kế ĐTN có các câu hỏi neo, các tham số, và các giá trị ước lượng của chúng (không kể sai số chọn mẫu) trong hai ĐTN có quan hệ tuyến tính, đó là

$$b_{Yc} = ab_{Xc} + \beta, a_{Yc} = a_{Xc} / \alpha,$$

trong đó  $b_{Xc}$  và  $b_{Yc}$  là các độ khó của các câu hỏi neo tương ứng trong các ĐTN X và ĐTN Y, khi các hằng số  $\alpha$  và  $\beta$  đã được xác định, các giá trị ước lượng tham số đối với mỗi câu hỏi trong ĐTN X có thể được đặt trên cùng thang đo với ĐTN Y. Các giá trị ước lượng tham số câu hỏi đối với các câu hỏi neo trong ĐTN X và trong ĐTN Y sẽ không như nhau (vì có sai số ước lượng), do đó sẽ được lấy trung bình.

Trong việc sử dụng cách thiết kế ĐTN có các câu hỏi neo, các câu hỏi neo nằm ở khoảng độ khó có thể chấp nhận đối với cả hai nhóm và thông thường số câu hỏi neo chiếm khoảng từ 20% đến 25% của số lượng câu hỏi trong các ĐTN.

### 3.3 Xác định các hằng số thiết lập thang đo

#### 3.3.1 Xác định các hằng số khi so bằng điểm thô

Theo Finch (2019), để chuyển điểm của TS từ ĐKT B sang ĐKT A ta sử dụng công thức sau đây:

$$x_A = a(x_B - c) + d, \quad (1)$$

với

$$a = \frac{\sqrt{\frac{2}{s_A^2} + b_{AZ1}^2 \left( \frac{2}{s_z^2} - \frac{2}{s_{z1}^2} \right)}}{\sqrt{\frac{2}{s_B^2} + b_{BZ2}^2 \left( \frac{2}{s_z^2} - \frac{2}{s_{z2}^2} \right)}},$$

$$c = \bar{x}_B + b_{BZ2}^2 (\bar{x}_z - \bar{x}_{z2}),$$

$$d = \bar{x}_A + b_{AZ1}^2 (\bar{x}_z - \bar{x}_{z1}),$$

Trong đó,

$x_A$  : điểm thô chuyển từ ĐKT B sang ĐKT A,

$x_B$  : điểm thô của ĐKT B,

$a, c, d$  : các tham số thang đo chung của hai ĐKT,

$s_A^2$  : phương sai điểm thô toàn câu hỏi của ĐKT A,

$s_B^2$  : phương sai điểm thô toàn câu hỏi của ĐKT B,

$\bar{x}_A$  : trung bình điểm thô toàn câu hỏi của ĐKT A,

$\bar{x}_B$  : trung bình điểm thô toàn câu hỏi của ĐKT B,

$s_z^2$  : phương sai điểm thô câu hỏi neo của ĐKT A và ĐKT B,

$s_{z1}^2$  : phương sai điểm thô câu hỏi neo của ĐKT A,

$s_{z2}^2$  : phương sai điểm thô câu hỏi neo của ĐKT B,

$\bar{x}_z$  : trung bình điểm thô câu hỏi neo của ĐKT A và ĐKT B,

$\bar{x}_{z1}$  : trung bình điểm thô câu hỏi neo của ĐKT A,

$\bar{x}_{z2}$  : trung bình điểm thô câu hỏi neo của ĐKT B,

$b_{AZ1}$  : hệ số góc của hàm hồi quy tuyến tính giữa điểm thô toàn câu hỏi và điểm thô câu hỏi neo trong ĐKT A,

$b_{BZ2}$  : hệ số góc của hàm hồi quy tuyến tính giữa điểm thô toàn câu hỏi và điểm thô câu hỏi neo trong ĐKT B.

#### 3.3.2 Xác định các hằng số khi so bằng điểm năng lực $\theta$ theo IRT

Theo Finch (2019), để chuyển điểm năng lực  $\theta$  theo IRT của TS từ ĐKT A sang ĐKT B ta sử dụng công thức sau đây:

$$\theta_B^* = A\theta_A + B, \quad (2)$$

Trong đó,

$\theta_B^*$  : điểm năng lực  $\theta$  theo IRT của TS từ ĐKT A sang ĐKT B,

$\theta_A$  : điểm năng lực  $\theta$  theo IRT của TS khi làm ĐKT A.

A, B là hai hằng số được xác định bởi công thức sau:

$$A = \frac{S_{Ab}}{S_{Bb}}; B = \bar{X}_{Ab} - A\bar{X}_{Bb},$$

với

$S_{Ab}$  : độ lệch chuẩn độ khó  $b$  (theo IRT) câu hỏi neo của ĐKT A,

$S_{Bb}$  : độ lệch chuẩn độ khó  $b$  (theo IRT) câu hỏi neo của ĐKT B,

$\bar{X}_{Ab}$  : trung bình độ khó  $b$  (theo IRT) câu hỏi neo của ĐKT A,

$\bar{X}_{Bb}$  : trung bình độ khó  $b$  (theo IRT) câu hỏi neo của ĐKT B.

**4 SO BẰNG HAI ĐỀ KIỂM TRA CHƯƠNG NGUYÊN HÀM – TÍCH PHÂN VÀ ỨNG DỤNG LỚP 12 SỬ DỤNG PHẦN MỀM IATA VÀ EXCEL**

**4.1 Quy trình biên soạn và so bằng ĐKT TNKQ**

Qua nghiên cứu các tài liệu tập huấn Kỹ thuật viết câu hỏi TNKQ, Công văn số 8773/BGDĐT-GDTrH ngày 30/12/2010 về việc hướng dẫn biên soạn đề kiểm tra, chúng tôi tiến hành so bằng ĐKT TNKQ chương Nguyên hàm – Tích phân và ứng dụng lớp 12 theo một quy trình như sau:

- Bước 1:** Xác định mục đích của đề kiểm tra
- Bước 2:** Thiết lập ma trận ĐKT
- Bước 3:** Biên soạn câu hỏi theo ma trận ĐKT
- Bước 4:** Xây dựng bảng đáp án và thang điểm
- Bước 5:** Kiểm tra, chỉnh sửa ĐKT trước khi so bằng

**Bước 6:** Chuyển điểm thô và điểm năng lực  $Q$  từ ĐKT này sang ĐKT kia

**Bước 7:** Kết luận về độ khó giữa hai ĐKT

**4.2 Ma trận đề của hai ĐKT**

Chúng tôi biên soạn và thẩm định hai ĐKT chương Nguyên hàm – Tích phân và ứng dụng, mỗi ĐKT gồm 30 câu hỏi TNKQ và thời gian làm bài là 54 phút (lấy theo mốc thời gian của Bộ Giáo dục và Đào tạo 1,8 phút/câu). Theo Bộ Giáo dục và Đào tạo (2010), ma trận chi tiết mô tả chuẩn kiến thức, kỹ năng, mục tiêu kiểm tra của hai ĐKT TNKQ được trình bày như sau:

**Bảng 1: Ma trận đề kiểm tra chương nguyên hàm – tích phân và ứng dụng**

Nhận thức Chủ đề	Nhận biết	Thông hiểu	Vận dụng		Tổng cộng
			Cấp độ thấp	Cấp độ cao	
<b>Chủ đề 1</b> <b>Nguyên hàm</b>	- Nhớ khái niệm nguyên hàm của một hàm số. - Biết các tính chất cơ bản của nguyên hàm.	- Tìm được nguyên hàm của một số hàm tương đối đơn giản dựa vào bảng nguyên hàm và cách tính nguyên hàm từng phần.	- Sử dụng được phương pháp đổi biến số (khi đã chỉ rõ cách đổi biến số và không đổi biến số quá một lần) để tính nguyên hàm. - Tìm một hàm số cụ thể nhờ xác định nguyên hàm.	- Giải quyết được bài toán thực tế	<b>8 câu</b> <b>26,7%</b> <b>[2,7 điểm]</b>
	<b>[2 câu]</b> <b>[0,7 điểm]</b>	<b>[3 câu]</b> <b>[1 điểm]</b>	<b>[2 câu]</b> <b>[0,7 điểm]</b>	<b>[1 câu]</b> <b>[0,3 điểm]</b>	
<b>Chủ đề 2</b> <b>Tích phân</b>	- Nhớ định nghĩa tích phân của hàm số liên tục bằng công thức Newton – Leibniz.	- Tính được tích phân của một số hàm tương đối đơn giản.	- Sử dụng phương pháp tính tích phân từng phần và phương pháp đổi biến số (khi		

	- Nhớ các tính chất của tích phân.		đã chỉ rõ cách đổi biến số và không đổi biến số quá một lần để tính tích phân của một hàm số.	<b>9 câu</b> <b>30%</b> <b>[3 điểm]</b>
	<b>[2 câu]</b> <b>[0,7 điểm]</b>	<b>[3 câu]</b> <b>[1 điểm]</b>	<b>[4 câu]</b> <b>[1,3 điểm]</b>	
<b>Chủ đề 3</b> <b>Ứng dụng của tích phân</b>	- Biết các công thức tính diện tích hình phẳng, thể tích vật thể, thể tích khối tròn xoay nhờ tích phân.	- Tính được diện tích, thể tích của một số hình giới hạn bởi các hàm số đơn giản.	- Tính được diện tích, thể tích một số hình phải xác định các cận.	- Giải quyết được bài toán thực tế bằng cách sử dụng tích phân.
	<b>[3 câu]</b> <b>[1 điểm]</b>	<b>[2 câu]</b> <b>[0,7 điểm]</b>	<b>[3 câu]</b> <b>[1 điểm]</b>	<b>13 câu</b> <b>43,3%</b> <b>[4,4 điểm]</b>
<b>Tổng tỉ lệ Điểm</b>	<b>7 câu</b> <b>23,3%</b> <b>[2,3 điểm]</b>	<b>8 câu</b> <b>26,7%</b> <b>[2,7 điểm]</b>	<b>9 câu</b> <b>30%</b> <b>[3 điểm]</b>	<b>6 câu</b> <b>20%</b> <b>[2 điểm]</b>
				<b>30 câu</b> <b>100%</b> <b>[10 điểm]</b>

Mỗi ĐKT gồm 30 câu hỏi TNKQ, trong đó có 6 câu hỏi neo (Câu 17, 18, 23, 25, 26, 30), trong mỗi câu hỏi gồm 4 phương án lựa chọn và chỉ có duy nhất một đáp án đúng. Các câu hỏi trong mỗi đề được sắp xếp theo mức độ nhận thức tăng dần từ nhận biết đến vận dụng (ở cấp độ). Trong mỗi mức độ, các câu hỏi được sắp xếp theo từng chủ đề (thứ tự như trong ma trận của ĐKT).

Trong mỗi câu hỏi, các phương án nhiễu được chúng tôi cân nhắc, dự đoán những sai lầm mà HS có thể gặp phải trong quá trình sử dụng kiến thức hoặc trong quá trình giải bài tập.

Toàn bộ quá trình thực nghiệm được thực hiện trên 544 HS thuộc 17 lớp khối 12 đang học tại 6 trường THPT ở các tỉnh Bạc Liêu, Sóc Trăng và Kiên Giang.

IATA (Item and Test Analysis) là một phần mềm được cung cấp miễn phí bởi Fernando Cartwright dùng để phân tích dữ liệu trong đánh giá giáo dục và tâm lý học.

Phần mềm IATA có thể tải về từ địa chỉ sau

<https://polymetrika.com/Downloads/IATA>

Kết quả phân tích các đề kiểm tra dựa trên phần mềm IATA cho các kết quả các câu hỏi trong hai đề kiểm tra A và B đều đạt yêu cầu về độ khó, độ phân biệt theo CTT và IRT. Độ tin cậy của đề A là 0,8 và đề B là 0,81.

Theo Linda (2008), độ tin cậy của các đề TNKQ từ 0.8 trở lên là các đề rất tốt; đây là cơ sở để tiến hành các bước tiếp theo trong quy trình so bằng hai đề kiểm tra này.

**4.3 Xác lập thang đo chung và so bằng**

Sử dụng phần mềm IATA sẽ cung cấp các tham số của các câu hỏi trong hai đề kiểm tra; kết hợp phần mềm Excel để tính các tham số trong bảng 2.

4.3.1 So bằng điểm thô: chuyển điểm thô của TS từ ĐKT B sang ĐKT A

**Bảng 2: Các tham số thang đo chung của hai ĐKT**

Các tham số	Kết quả	Các tham số	Kết quả
$S_A^2$	378,30	$\bar{x}_A$	54,25
$S_B^2$	398,80	$\bar{x}_B$	52,37
$S_{Z1}^2$	881,99	$\bar{x}_{Z1}$	46,56
$S_{Z2}^2$	917,35	$\bar{x}_{Z2}$	43,93
$S_Z^2$	899,75	$\bar{x}_Z$	45,25
$b_{AZ1}$	0,52	$b_{BZ2}$	0,53
a	0,98	c	52,75
d	53,88		

thay  $a = 0,98, c = 52,75, d = 53,88$  vào công thức  $x_A = a(x_B - c) + d$  (1) thu được kết quả

chuyển điểm thô (%) của TS từ ĐKT B sang ĐKT A của TS như sau:

**Bảng 3: Kết quả chuyển điểm thô của TS từ ĐKT B sang ĐKT A**

ID	Điểm thô ĐKT B (%)	Điểm thô chuyển từ ĐKT B sang ĐKT A (%)
2001	23,33	24,85
2002	33,33	34,72
2003	26,67	28,14
2004	83,33	84,05
2005	40	41,29
.....	.....	.....
2268	40	41,29
2269	40	41,29
2270	36,67	38,01
2271	66,67	67,61

Từ kết quả chuyển điểm thô từ ĐKT B sang ĐKT A, hầu hết điểm thô (%) tăng khi chuyển điểm thô từ ĐKT B sang ĐKT A. Ví dụ, TS có ID 2019 điểm thô khi làm ĐKT B là 80 điểm và khi chuyển điểm sang ĐKT A là 80,76 điểm tăng lên 0,76 điểm, TS có ID 2254 điểm thô khi làm ĐKT B là 40 điểm và khi chuyển điểm sang ĐKT A là 41,29, tăng lên 1,29 điểm, điều đó chứng tỏ ĐKT B khó hơn ĐKT A. Tuy nhiên, độ khó chênh lệch là không lớn.

4.3.2 So sánh điểm năng lực  $Q$  của TS từ ĐKT A sang ĐKT B

**Bảng 4: Độ khó b các câu hỏi neo và các tham số thang đo chung của hai ĐKT**

Câu hỏi neo	Độ khó b của ĐKT A	Độ khó b của ĐKT B
Câu 17	-0,13	-0,02
Câu 18	0,37	0,1
Câu 23	0,37	0,55
Câu 25	-0,48	0,03
Câu 26	-0,05	0,08
Câu 30	0,71	0,53
Trung bình	$\bar{X}_{Aa} = 0,13$	$\bar{X}_{Bb} = 0,21$
Độ lệch chuẩn	$S_{Ab} = 0,42$	$S_{Ab} = 0,25$
Hàng số	$A = \frac{S_{Ab}}{S_{Bb}} = \frac{0,42}{0,25} = 1,66$	
	$B = \bar{X}_{Ab} - A\bar{X}_{Bb} = -0,22$	

thay  $A = 1,66, B = -0,22$  vào công thức  $\theta_B^* = A\theta_A + B (2)$  thu được kết quả chuyển điểm năng lực  $\theta$  của TS từ ĐKT A sang ĐKT như sau:

**Bảng 5: Kết quả điểm năng lực  $\theta$  của TS chuyển từ ĐKT A sang ĐKT B**

ID	Điểm năng lực $\theta$ của ĐKT A	Điểm năng lực $\theta$ chuyển từ ĐKT A sang ĐKT B
1001	1,15	1,70
1002	0,04	-0,15
1003	0,24	0,18
1004	0,76	1,05
1005	0,36	0,38
.....	.....	.....
1268	-2,39	-4,21
1269	-1,76	-3,15
1270	-0,56	-1,16
1271	-2,35	-4,14
1272	-0,95	-1,80

Từ kết quả chuyển điểm năng lực  $\theta$  của TS theo IRT từ ĐKT A sang ĐKT B, điểm năng lực  $\theta$  hầu hết là giảm khi chuyển từ ĐKT A sang ĐKT B. Ví dụ, TS có ID 1002 điểm năng lực  $\theta$  khi làm ĐKT A là 0,04 và khi chuyển điểm sang ĐKT B là -0,15, TS có ID 1272 điểm năng lực  $\theta$  khi làm ĐKT A là -0,95 và khi chuyển điểm sang ĐKT B là -1,8 điều đó chứng tỏ ĐKT B khó hơn ĐKT A.

**5 KẾT LUẬN**

Từ kết quả nghiên cứu này, chúng ta thấy rằng việc so sánh hai ĐKT trước tiên đòi hỏi hai ĐKT này phải được biên soạn tốt và độ tin cậy của hai ĐKT phải ở mức độ cho phép và tương đương nhau. Bên cạnh đó, để khai thác tối đa và hiệu quả hết các tính năng của phần mềm IATA cũng như nắm được nhiều thông tin nâng cao hơn về ĐKT, cần có những nghiên cứu chuyên sâu, chi tiết và quy mô hơn.

So sánh hai ĐKT với sự hỗ trợ của phần mềm IATA và Excel là một phương pháp tương đối mới (ở Việt Nam) để đánh giá về độ khó của hai ĐKT; so sánh là việc thiết lập thang đo chung của hai ĐKT, từ đó tạo sự công bằng cho HS khi làm các ĐKT khác nhau và cũng là công cụ tốt giúp xác định và so sánh năng lực của HS ở trường THPT. Ngoài ra, nghiên cứu này còn có thể phát triển lên so sánh ba hay bốn ĐKT. Phần mềm IATA hoặc những phần mềm tương tự đang được ứng dụng rộng rãi trong việc so sánh ĐKT, góp phần cải cách kiểm tra, đánh giá trong giáo dục nói chung, và nâng cao chất lượng giáo dục môn Toán nói riêng.

**TÀI LIỆU THAM KHẢO**

Bộ Giáo dục và Đào tạo, 2010. Hướng dẫn biên soạn đề kiểm tra (Kèm theo công văn số 8773/BGDĐT-GDTrH), Hà Nội.

- Bộ Giáo dục và Đào tạo, 2017. Hướng dẫn thực hiện Quy chế thi THPT quốc gia và xét công nhận tốt nghiệp THPT năm 2017. Công văn số Số 417/BGDĐT KTKĐCLGD, Hà Nội.
- Dương Thiệu Tông, 2005. Trắc nghiệm và đo lường thành quả học tập. NXB Khoa học xã hội, TP. Hồ Chí Minh, 545 trang.
- Lâm Quang Thiệp, 2011. Đo lường trong giáo dục – Lý thuyết và ứng dụng. NXB Đại học Sư phạm, Hà Nội, 291 trang.

- Ayala, R., 2009. The Theory and Practice of Item Response Theory. The Guilford Press, New York, USA, 448 pages.
- Finch, W.H. and French, B.F., 2019. Educational and Psychological Measurement. Routledge, New York and London, 469 pages.
- Linda, C. and James, A., 2008. Introduction to Classical and Modern Test Theory. Cengage Learning, USA, 527 pages.
- Thomas, M.H., 2004. Developing and Validating Multiple – Choice Test Items. Lawrence Erlbaum Associates, London, 306 pages.