

# Phân tích câu hỏi và đề thi theo lí thuyết khảo thí cổ điển bằng các phần mềm Excel và IATA

Vũ Đó Long<sup>1</sup>, Nguyễn Văn Dũng<sup>2</sup>,  
Vũ Thị Thảo<sup>3</sup>, Nguyễn Thị Mỹ Linh<sup>4</sup>

Email: longvd@vnu.edu.vn  
Email: dungnvtkt@vnu.edu.vn  
Email: thaovttkd@vnu.edu.vn  
Email: lhmimtkd@vnu.edu.vn

Trung tâm Khảo thí, Đại học Quốc gia Hà Nội  
44 Xuân Thủy, Cầu Giấy, Hà Nội, Việt Nam

**TÓM TẮT:** Khi đánh giá câu hỏi và đề thi theo lí thuyết khảo thí cổ điển, người thực hiện thường dùng phần mềm để phân tích các đại lượng đánh giá điểm thi như điểm trung bình, trung vị, phương sai, độ lệch chuẩn và các đại lượng đánh giá câu hỏi, để thi như độ khó, độ phân biệt, độ tin cậy, hệ số tương quan (Point-biserial). Bài viết trình bày ý nghĩa của một số đại lượng đo lường cơ bản trong lí thuyết khảo thí cổ điển, phân tích bài làm trắc nghiệm khách quan bằng phương pháp sử dụng công thức với sự hỗ trợ của phần mềm Excel và so sánh kết quả thu được với kết quả tính toán bằng phần mềm phân tích miễn phí IATA. Từ đó, có những kết luận về căn cứ tính toán và độ chính xác của phần mềm IATA.

**TỪ KHÓA:** Phân tích câu hỏi và đề thi; lí thuyết khảo thí cổ điển; EXCEL; IATA.

→ Nhận bài 12/6/2019 → Nhận kết quả phản biện và chỉnh sửa 18/7/2019 → Duyệt đăng 25/8/2019.

## 1. Đặt vấn đề

Phân tích câu hỏi và đề thi là bước quan trọng để đánh giá chất lượng ngân hàng câu hỏi và đề thi. Kết quả phân tích là cơ sở định lượng giúp người đánh giá có căn cứ thực tiễn để điều chỉnh, tác động một cách phù hợp nhằm hạn chế được sự sai lệch định tính khi xây dựng câu hỏi, đề thi và đảm bảo được sự công bằng, chính xác trong đánh giá. Việc tính toán các chỉ số đo lường thường được hỗ trợ bởi các phần mềm. Tuy nhiên, đa số các phần mềm khảo thí chuyên nghiệp thường có chi phí bản quyền cao trong khi các phần mềm miễn phí còn gây nhiều khó khăn trong cách sử dụng, cách xác định ý nghĩa kết quả thống kê và đôi khi không tổng nhất với nhau về kết quả và căn cứ tính toán. Nhằm xác định rõ ý nghĩa của các chỉ số đo lường cơ bản bằng quá trình tính toán, nhóm tác giả tiến hành phân tích câu hỏi, đề thi bằng phương pháp sử dụng công thức của lí thuyết khảo thí cổ điển với sự hỗ trợ của phần mềm Excel và so sánh với kết quả phân tích bằng phần mềm phân tích miễn phí IATA (Item And Test Analysis) version 7.4. Kết quả tính toán cũng là cơ sở để đánh giá sự chính xác và kiểm định độ tin cậy của phần mềm IATA nói riêng và các phần mềm phân tích miễn phí nói chung.

## 2. Nội dung nghiên cứu

**2.1. Các chỉ số đo lường đặc trưng cơ bản của lí thuyết khảo thí cổ điển sử dụng trong phân tích đề thi và câu hỏi thi**

**2.1.1. Điểm trung bình, điểm trung vị và điểm phổ biến**

Điểm trung bình (Mean) của bài thi được tính dựa trên điểm số của các thí sinh dự thi bài thi đó. Điểm trung bình của bài thi được tính theo công thức sau:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{N} \sum_{i=1}^N n_i x_i$$

Trong đó:  $\bar{x}$  là giá trị trung bình điểm thi của bài thi,  $x_i$

là điểm của thí sinh thứ ( $i$ );

$N$  là tổng số thí sinh;  $n_i$  là tần số-số thí sinh được  $x_i$  điểm. Điểm trung vị (Median) là điểm đứng giữa tập số liệu điểm đã được sắp xếp theo thứ tự từ bé đến lớn, chia dãy số điểm thi đó làm 2 phần bằng nhau về số điểm. Khi điểm thi của các thí sinh giữa các nhóm có sự chênh lệch lớn rõ rệt thì điểm trung bình không đại diện cho mức điểm trung bình của đa số các thí sinh, khi đó người ta kết hợp đánh giá bằng điểm trung vị. Điểm phổ biến (Mode) là điểm có tần số xuất hiện nhiều nhất trong số các kết quả điểm thi thu được.

### 2.1.2. Phương sai, độ lệch chuẩn

Phương sai ( $\sigma^2$  hoặc  $S^2$ ), là đại lượng biểu diễn độ phân tán của tập số liệu kết quả đo đối với giá trị trung bình [1]. Phương sai được tính theo công thức sau:

$$\sigma^2 = S^2 = \frac{1}{N'} \sum_{i=1}^N (x_i - \bar{x})^2$$

Với:  $N' = N$  khi  $N > 30$ ;  $N' = N - 1$  khi  $N < 30$ .

Trong đó:  $N$  là tổng số thí sinh;  $x_i$  là điểm của thí sinh thứ ( $i$ );  $\bar{x}$  là điểm trung bình toàn bài thi.

Độ lệch chuẩn là đại lượng thể hiện độ phân tán của tập số liệu kết quả đo đối với giá trị trung bình.

Độ lệch chuẩn hay độ lệch tiêu chuẩn của một tập kết quả thi là giá trị căn bậc hai trị số phương sai của nó:  $\sigma = \sqrt{\sigma^2}$

hoặc  $S = \sqrt{S^2}$

### 2.1.3. Độ khó, độ phân biệt

Độ khó ( $p$ ) của một câu hỏi thi được tính bằng tỉ lệ giữa số thí sinh trả lời đúng với tổng số thí sinh tham gia trả lời câu hỏi thi đó. Theo đó, giá trị của  $p$  càng thấp thì độ khó

càng cao và ngược lại,  $0 \leq p \leq 1$ .

Độ phân biệt của câu hỏi thì là đại lượng thể hiện mức độ khác nhau về kết quả trả lời một câu hỏi của nhóm năng lực cao và nhóm năng lực thấp.

$$D = \frac{N_H - N_L}{n}$$

Trong đó:  $N_H$  là số thí sinh thuộc nhóm năng lực cao trả lời đúng câu hỏi;  $N_L$  là số thí sinh thuộc nhóm năng lực thấp trả lời đúng câu hỏi;  $n$  là số thí sinh của nhóm năng lực cao hoặc nhóm năng lực thấp;  $-1 \leq D \leq 1$ .

Thông thường, nhóm năng lực cao lấy trong khoảng 27% thí sinh có điểm cao nhất, nhóm năng lực thấp là nhóm 27% thí sinh có điểm thấp nhất.

### 2.1.4. Hệ số tương quan

Hệ số tương quan ( $r$ ) là một chỉ số thống kê đo lường mối liên hệ tương quan giữa hai biến.

Công thức tính hệ số tương quan:

$$r = \frac{S_{xy}}{S_x S_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Hệ số *Point-biserial* của câu hỏi trong bài test là tương quan giữa điểm của câu hỏi (0 hoặc 1) với điểm tổng thể của từng thí sinh. Chẳng hạn, xét câu hỏi thứ ( $m$ ), có  $N$  thí sinh trả lời, ta được các dãy số liệu sau:

X	$x_{m1} = \{0,1\}$	$x_{m2} = \{0,1\}$	...	$x_{mn} = \{0,1\}$
Y	$y_1$	$y_2$	...	$y_N$

Với  $x_{mi}$  là điểm trả lời câu hỏi thứ ( $m$ ) của học sinh thứ ( $i$ ), giá trị  $x_{mi} = 0$  (sai) hoặc  $x_{mi} = 1$  (đúng), giá trị  $y_i$  là điểm toàn bài thi của học sinh thứ ( $i$ ),  $\forall i = 1, 2, \dots, N$ .

Hiển nhiên,  $y_i = x_{i1} + x_{i2} + \dots + x_{in}$ ;  $n$  là tổng số câu hỏi đề thi.

Đặt  $\bar{x}_m = (x_{m1} + x_{m2} + \dots + x_{mn}) / N$  là điểm trung bình câu hỏi thứ ( $m$ ) của tất cả  $N$  thí sinh;  $\bar{y} = (y_1 + y_2 + \dots + y_N) / N$  là điểm trung bình toàn bài thi của tất cả  $N$  thí sinh.

Kí hiệu hệ số *Point-biserial* của câu hỏi thứ ( $m$ ) là  $R_{pb}$ . Ta xem  $R_{pb}$  như là hệ số tương quan hai dãy X và Y thu được trong bảng trên.

$$R_{pb} = \frac{\sum (x_{mi} - \bar{x}_m)(y_i - \bar{y})}{\sqrt{(\sum (x_{mi} - \bar{x}_m)^2)(\sum (y_i - \bar{y})^2)}}$$

Miền giá trị thu được  $-1 \leq R_{pb} \leq 1$ . Độ tin cậy của điểm thi tăng lên khi câu hỏi thì và bài thi có mối tương quan chặt chẽ.

### 2.1.5. Độ tin cậy của đề thi

Độ tin cậy (reliability) là đại lượng thể hiện sự chính xác của phép đo lường. Độ tin cậy của kết quả thi tốt thì hệ số tương quan giữa hai tập hợp điểm của cùng một nhóm thí sinh cao nghĩa là với cùng một nhóm thí sinh tham gia

hai bài thi cho kết quả điểm số là tương đương với nhau [2].

Công thức tính độ tin cậy Kuder - Richardson [3]:

$$KR20 = \delta = \frac{n}{n-1} \left[ 1 - \frac{1}{\sigma^2} \sum_{i=1}^n p_i(1-p_i) \right]$$

trong đó,  $p_i$  là tỉ lệ số thí sinh trả lời đúng câu hỏi thứ  $i$ ;  $\sigma^2$  là phương sai chung của cả bài test;  $n$  là số câu hỏi của bài test

Công thức phân đôi:

$$\delta = \frac{2r}{1+r}$$

Trong đó,  $r$  là độ tin cậy của nửa bài test. Bài test thường được phân đôi theo hai dạng bao gồm: Nửa trên/nửa dưới hoặc nửa chẵn/nửa lẻ. Giá trị  $r$  được xác định như tương quan của hai dãy nửa trên/nửa dưới hoặc nửa chẵn/nửa lẻ.

Công thức dự báo của Spearman - Brown:

$$\delta = \frac{kr}{1+(k-1)r}$$

Trong đó,  $r$  là độ tin cậy của bài test;  $\delta$  là độ tin cậy của bài test được kéo dài  $k$  lần.

Công thức phân đôi là tương hợp đặc biệt của công thức Spearman - Brown khi  $k=2$ .

Khi đề cập đến độ tin cậy, người ta thường tính hệ số tin cậy  $\alpha$ -Cronbach:

$$\alpha = \frac{n}{n-1} \left( 1 - \frac{1}{\sigma^2} \sum_{i=1}^n \sigma_i^2 \right)$$

trong đó,  $\sigma_i^2$  là phương sai của dãy kết quả trả lời câu hỏi thứ ( $i$ );  $\sigma^2$  là phương sai chung của kết quả cả bài test;  $n$  là số câu hỏi trong bài test.

Giá trị hệ số  $\alpha$ -Cronbach bằng giá trị KR20 với bài test trắc nghiệm Đúng - Sai.

### 2.1.6. Điểm Z-score

Trong đo lường điểm thi, các phép đo điểm thô còn có sai số và độ lệch nhất định so với điểm thực. Chính vì vậy, việc dùng điểm thô để kết luận sẽ có độ tin cậy thấp và trong tính toán hay dùng điểm Z - score để biểu diễn. Với  $\sigma$  là độ lệch chuẩn của biến ngẫu nhiên, thì Z - score được tính theo công thức dưới đây:

$$Z_i = \frac{x_i - \bar{x}}{\sigma}$$

Với  $x_i$  là giá trị điểm của thí sinh thứ ( $i$ ),  $\bar{x}$  là điểm trung bình của các thí sinh.

### 2.1.7. Độ lệch, độ nhọn

$$Skewness = \frac{N}{(N-1)(N-2)} \cdot \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \bar{x})^3$$

Trong đó:  $N$  là tổng số thí sinh  
Phân bố chuẩn sẽ có Skewness xấp xỉ 0, khi Skewness lớn hơn 0 thì đường cong phân bố lệch trái, khi Skewness nhỏ hơn 0 thì đường cong phân bố lệch phải.

$$Kurtosis = \frac{N(N+1)}{(N-1)(N-2)(N-3)} \cdot \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \bar{x})^4$$

Một phân bố chuẩn sẽ có Kurtosis xấp xỉ 3, khi Kurtosis m hơn 3 thì đường cong phân bố nhọn, khi Kurtosis nhỏ hơn 3 thì đường cong phân bố rộng.

## 2.2. Đánh giá câu hỏi trắc nghiệm khách quan bằng phần mềm IATA

### 2.2.1. Giới thiệu về phần mềm IATA (ITEM AND TEST ANALYSIS)

Phần mềm IATA là phần mềm đánh giá phân tích câu hỏi, đề thi theo lý thuyết đáp ứng câu hỏi (Item Response theory – IRT). Đây là phần mềm miễn phí, có tính năng đơn giản, dễ sử dụng nên được sử dụng rộng rãi trong phân tích câu hỏi thi. Để sử dụng các chức năng và đọc kết quả thông qua các chỉ số quy ước của phần mềm, người dùng có thể tải phần mềm, cài đặt và sử dụng theo hướng dẫn trên trang chủ <https://polymetrika.com/Downloads/IATA> (xem hình 1) [4].

Để sử dụng phần mềm với đầy đủ các chức năng, người dùng không nhất thiết phải đăng kí tài khoản. Phần mềm có thể lựa chọn tiếng Việt là một trong 8 ngôn ngữ được tích hợp sẵn. Phần mềm có các chức năng nổi bật như: Phân tích dữ liệu câu trả lời của thí sinh, phân tích dữ liệu câu trả lời có liên kết, liên kết dữ liệu (linking items), lựa chọn câu hỏi ưu, xây dựng và gán các tiêu chuẩn thành tích.

### 2.2.2. Hướng dẫn phân tích đề thi và câu hỏi thi trắc nghiệm khách quan bốn lựa chọn bằng phần mềm IATA

**Định dạng dữ liệu:** Theo Fernando Cartwright trong hướng dẫn sử dụng phần mềm IATA [4] thì phần mềm IATA chấp nhận dữ liệu đầu vào dạng bảng lưu dưới dạng file có uoi dạng \*.xls/\*.xlsx/\*.csv (file Excel), \*.sav (file SPSS) hoặc \*.txt.

**Nhập dữ liệu:** Chọn chức năng “Phân tích dữ liệu câu hỏi”, tiến hành nhập bằng dữ liệu. Bảng 1 dưới đây là một ví dụ trong các dạng dữ liệu đầu vào có chứa thông tin là các phương án trả lời của thí sinh khi làm bài thi trắc nghiệm bốn lựa chọn.

Bảng 1: Dữ liệu câu trả lời của thí sinh

Ho	Tên	SBD	Câu 1	Câu 2	Câu 3	Điểm
Nguyen	Dung	12016	A	B	D	6.0
Van						
Vu Thanh	Vuong	12005	C	D	A	7.5

Tiếp tục nhập dữ liệu đáp án của các câu hỏi (xem Bảng 2) để chấm điểm và tính toán các chỉ số của câu hỏi, bài thi và thống kê theo từng thí sinh tham gia trả lời.

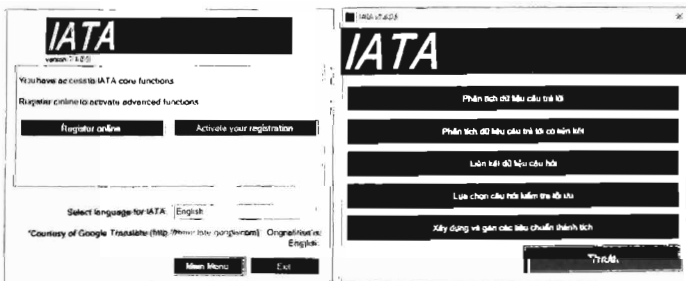
Bảng 2: Dữ liệu đáp án của các câu hỏi (item)

Questions	Key	Level	Content
Câu 1	B	1	NL1
Câu 2	C	2	NL2
Câu 3	A	3	NL3

Sau khi nhập dữ liệu, người dùng cũng có thể hiệu chỉnh số liệu trực tiếp trên phần mềm.

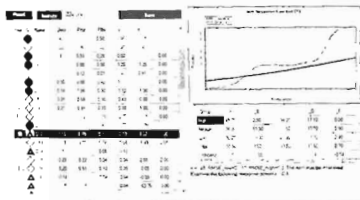
**Kết quả tính toán:** Phần mềm IATA có thể tính toán được một số đại lượng đo lường về chất lượng của câu hỏi và đề thi bao gồm: Giá trị trung bình, trung vị, độ lệch chuẩn, phương sai, độ tin cậy của kết quả thi, độ khó, độ phân biệt, hệ số đoàn mớ, hệ số tương quan của câu hỏi với với đề thi, điểm cao nhất, điểm thấp nhất, tỉ lệ chọn các phương án của ba nhóm thí sinh có năng lực thấp, năng lực trung bình và năng lực cao... Các giá trị này được tính theo các lý thuyết khảo thí cổ điển, lý thuyết khảo thí hiện đại và theo các loại điểm như điểm phần trăm, điểm Zscore.

Đối với mỗi câu hỏi, dựa vào các chỉ số tính toán được phần mềm đưa ra cảnh báo dưới dạng hình ảnh, màu sắc và ghi chú, giúp người sử dụng có thể phát hiện nhanh các câu hỏi có vấn đề cần xem xét. Các câu có chỉ số thông kê tốt có màu xanh, các câu hỏi có vấn đề hoặc cần xem xét có màu đỏ hoặc vàng tùy theo mức độ. Ngoài ra, phần mềm còn cho



Hình 1: Giao diện khi khởi động và các chức năng chính của phần mềm IATA

ta các biểu đồ về phân bố điểm thi, đường phân bố năng lực của thí sinh và độ khó của câu hỏi (xem Hình 2).

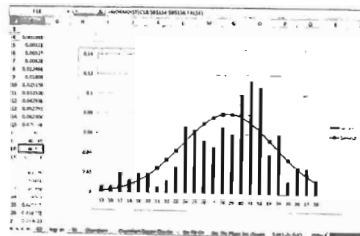


Hình 2: Minh họa một số kết quả phân tích câu hỏi bằng phần mềm IATA

Trong Hình 2, câu hỏi số 1 (C1) có các chỉ số đo lường tốt, phù hợp với tiêu chuẩn chung và được tô màu xanh. Câu hỏi số 15 (C15) có độ khó cao (PVal) = 0.22 nhưng độ phân biệt thấp (Discr = 0.28) nên được cảnh báo mức màu vàng. Câu số 12 (C12) có cảnh báo màu đỏ. Ta thấy đây là câu hỏi có độ khó PVal = 0.39 ở mức trên trung bình, tuy nhiên độ phân biệt Discr = 0.15 và mức tương quan Pbis=0.17 là khá thấp. Đường biểu diễn tương quan giữa năng lực của thí sinh và độ khó của câu hỏi có độ dốc không đều, cách xa đường ki vông. Do vậy, đây là câu hỏi chưa tốt cần xem xét lại. Như vậy, dựa vào màu sắc của các câu hỏi, có thể tìm nhanh các câu hỏi có vấn đề để xem xét, điều chỉnh và sàng lọc.

### 2.3. So sánh kết quả phân tích bằng phương pháp tính toán theo công thức sử dụng phần mềm Excel và phương pháp sử dụng phần mềm IATA

Thực hiện phân tích đề thi gồm 50 câu hỏi thi trắc nghiệm kết quả bốn lựa chọn bằng phương pháp lý thuyết sử dụng công cụ hỗ trợ là phần mềm Excel và phương pháp sử dụng phần mềm IATA để so sánh (xem Hình 3). Mẫu dữ liệu là bài làm của 148 thí sinh với các lựa chọn A, B, C, D của mỗi câu hỏi. Điểm được chấm theo thang điểm 100. Kết



Hình 3: Đồ thị phân bố điểm thi vẽ bằng phần mềm Microsoft Excel

quả tính toán được phân loại và so sánh nhằm làm rõ hơn về ý nghĩa của các chỉ số và sự sai khác trong tính toán được thực hiện bằng hai phương pháp trên.

#### 2.3.1. So sánh các chỉ số đo lường về đề thi

Bảng 3: So sánh các chỉ số đo lường điểm bài thi

Các chỉ số đo lường về đề thi	Kết quả tính toán bằng phần mềm		Độ lệch kết quả (3) = (1) - (2)
	Excel (1)	IATA (2)	
Mean	57.108	57.110	0.002
Median	60.00	58	2
Min	30.00	30.00	0
Max	76.00	76.00	0
Độ lệch chuẩn (stDev)	9.898	9.890	0.008
Độ tin cậy (Kf20)	0.645	0.660	0.015
Độ lệch (Skewness)	-0.565	-0.570	0.050
Độ nhọn (Ex. Kurtosis)	0.016	-0.070	0.054

**Nhận xét:** Như vậy, các chỉ số đo lường về đề thi được tính bằng hai phương pháp có độ chênh lệch không lớn, xấp xỉ nhau (xem Bảng 3).

#### 2.3.2. So sánh các chỉ số đo lường về câu hỏi thi

Bảng 4: So sánh các chỉ số đo lường câu hỏi thi

Câu hỏi	Độ khó		Độ phân biệt		Hệ số tương quan	
	Excel	IATA	Excel	IATA	Excel	IATA
Câu 1	0.80	0.80	0.43	0.42	0.50	0.50
Câu 2	0.86	0.86	0.18	0.19	0.31	0.31
Câu 3	0.93	0.93	0.18	0.18	0.28	0.28
...	...	...	...	...	...	...
Câu 49	0.80	0.80	0.33	0.31	0.38	0.38
Câu 50	0.43	0.43	0.33	0.35	0.29	0.29

**Nhận xét:** Từ bảng so sánh ta thấy các giá trị độ khó được tính bằng hai phương pháp trên các công cụ EXCEL và IATA là như nhau, độ phân biệt có mức chênh lệch từ 0.01 đến 0.05 và một số hệ số tương quan có mức chênh lệch là 0.01 đơn vị (xem Bảng 4). Như vậy, kết quả tính toán bằng phần mềm đa số có kết quả chính xác như phương pháp tính toán bằng lý thuyết sử dụng công thức.

### 3. Kết luận và đề xuất

Việc sử dụng phần mềm miễn phí như IATA để phân tích

nhiệm sai lệch nhất định do làm tròn số hay các nguyên nhân khác nhau. Tuy mức chênh lệch không cao nhưng cũng có thể gây ra những kết luận không chính xác. Khi phân tích cần chú ý không máy móc căn cứ quá sát vào các số để đối sánh mà lí thuyết khảo thí đưa ra. Cần xem xét kĩ các đại lượng có giá trị tiệm cận các mốc chia khoảng như  $\mu$  và có thể tính toán lại bằng lí thuyết để so sánh khi cần. Bài viết đã mô tả các cơ sở của thuật toán lập trình phân tích khảo thí, câu hỏi thi trên EXCEL. Kết quả so sánh các chỉ số

và phần mềm IATA là gần như trùng nhau, điều này cho thấy độ tin cậy của phương pháp đề xuất.

Cách tính này đã được Trung tâm khảo thí Trường Đại học Quốc gia Hà Nội dùng để viết các thuật toán xây dựng phần mềm khảo thí và sử dụng trong giảng dạy lớp tập huấn về kiểm tra đánh giá. Các bước tính toán trong môi trường EXCEL còn làm rõ hơn ý nghĩa của các đại lượng thu được, giúp người phân tích đưa ra được những nhận xét, đánh giá chính xác hơn về chất lượng của câu hỏi và đề thi.

#### Tài liệu tham khảo

- [1] Lê Đức Ngọc, (2018), *Thống kê ứng dụng trong đo lường và đánh giá giáo dục*. Tài liệu giảng dạy.
- [2] Lâm Quang Thiệp, (2010), *Đo lường trong giáo dục-Lí thuyết và ứng dụng*. NXB Đại học Quốc gia Hà Nội
- [3] Phạm Xuân Thanh, (2007), *Lí thuyết đánh giá (tài liệu giảng dạy lớp thực sĩ đo lường đánh giá khóa 1, 2)*.
- [4] Fernando Cartwright, (2013), *IATA Manual: Item and Test Analysis (IATA)*.

## ITEM ANALYSIS IN CLASSICAL TEST THEORY USING MS EXCEL AND IATA SOFTWARES

Do Long<sup>1</sup>, Nguyen Van Dung<sup>2</sup>,  
Thi Thao<sup>3</sup>, Nguyen Thi My Linh<sup>4</sup>

mail: longvd@vnu.edu.vn  
mail: dungnvtdk@vnu.edu.vn  
mail: thaovttdk@vnu.edu.vn  
mail: linhntmtdk@vnu.edu.vn

Center for Educational Testing (CET)  
Xuan Thuy, Cau Giay, Hanoi, Vietnam

**ABSTRACT:** *The work has briefly overviewed the concepts and the meaning of measurements of the Classical Test Theory (CTT) and the relationships between statistical values such as mean, median, variance, standard deviation with corresponding values in item analysis such as item difficulty, item discrimination, test reliability, point-biserial coefficient, respectively. To evaluate algorithm and to obtain the accuracy of IATA software, the numerical values of measurements computed in IATA has been compared with corresponding ones calculated in MS Excel using original formulae of the Classical Test Theory.*

**KEYWORDS:** Item analysis; test analysis; Classical Test Theory; MS EXCEL, IATA.