

# PHÂN TÍCH CÂU HỎI TRẮC NGHIỆM KHÁCH QUAN DÙNG ĐÁNH GIÁ NĂNG LỰC VẬT LÝ CỦA HỌC SINH TRONG DẠY HỌC CHƯƠNG “SÓNG ÁNH SÁNG” – VẬT LÝ 12 BẰNG PHẦN MỀM QUEST/CONQUEST

## ANALYZING OBJECTIVE MULTIPLE CHOICE QUESTIONS FOR ASSESSING THE PHYSICS COMPETENCIES OF STUDENTS IN TEACHING THE “LIGHT WAVES” CHAPTER OF PHYSICS GRADE 12 USING QUEST/CONQUEST SOFTWARE

Nguyễn Bảo Hoàng Thanh<sup>1\*</sup>, Nguyễn Thị Thảo<sup>2</sup>

<sup>1</sup>Trường Đại học Sư phạm - Đại học Đà Nẵng

<sup>2</sup>Trường THPT Chu Văn An - Ninh Thuận

\*Tác giả liên hệ: nguyebaohoangthanh@gmail.com

(Nhận bài: 09/11/2022; Chấp nhận đăng: 07/3/2023)

**Tóm tắt** - Xu hướng mới trong đánh giá giáo dục hiện nay là sự đánh giá dựa trên năng lực học sinh (HS). Một trong những phương pháp đổi mới hình thức kiểm tra đánh giá là bổ sung thêm dạng câu hỏi trắc nghiệm (TN) khách quan trong bài kiểm tra đánh giá năng lực của HS. Ưu điểm nổi bật của phương pháp TN là có thể dễ dàng sử dụng nhiều phần mềm chuyên dụng để phân tích các câu hỏi TN, đề TN một cách nhanh chóng và chính xác như xác định độ phân biệt, độ khó, độ tin cậy, mức độ phù hợp của đề TN với mô hình lý thuyết IRT (Item Response Theory) cũng như các yếu tố về năng lực của HS và mối quan hệ giữa chúng có thể được mô tả dưới dạng văn bản và đồ thị. Kết quả phân tích và đánh giá câu hỏi TN khách quan dùng đánh giá năng lực vật lý của HS trong dạy học chương Sóng ánh sáng Vật lý 12 theo lý thuyết IRT được trình bày chi tiết trong bài báo này, với sự hỗ trợ của phần mềm Quest/Conquest.

**Từ khóa** - Lý thuyết ứng đáp câu hỏi; phần mềm Quest/Conquest; đánh giá năng lực; câu hỏi trắc nghiệm khách quan; năng lực vật lý.

### 1. Đặt vấn đề

Hiện nay, đổi mới căn bản toàn diện giáo dục, đào tạo đã và đang được ngành giáo dục triển khai thực hiện và xem đó là một nhiệm vụ cấp thiết. Trong quá trình đổi mới nền giáo dục với mục tiêu nâng cao chất lượng đào tạo, vai trò của kiểm tra đánh giá đã được xác nhận là một chiến lược và chính sách giáo dục quốc gia. Kiểm tra đánh giá đóng một vai trò quan trọng đặc biệt trong quá trình giáo dục và dạy học bằng cách cung cấp thông tin phản hồi cần thiết cho giáo viên, HS và nhà quản lý để điều chỉnh các thành phần còn lại của quá trình dạy học. Hiện nay, xu hướng mới của kiểm tra, đánh giá là kiểm tra, đánh giá theo năng lực. Kiểm tra, đánh giá là học tập chú trọng khả năng vận dụng sáng tạo tri thức trong những tình huống ứng dụng khác nhau trong thực tiễn. Việc kiểm tra, đánh giá năng lực nhằm hỗ trợ giáo viên có thông tin về kết quả học tập của HS, từ đó điều chỉnh phương pháp dạy học; Đồng thời giúp HS cải thiện hoạt động học tập; Ngoài ra kiểm tra, đánh giá còn hỗ trợ giáo viên và nhà trường trong việc xác nhận kết quả học tập của HS. Nhằm đáp ứng yêu cầu đổi mới trong việc kiểm tra đánh giá kết quả học tập của HS theo năng lực, giáo viên có thể sử dụng linh hoạt các hình thức đánh giá như quan sát, vấn đáp, tự luận, sản phẩm

**Abstract** - The new trend in the educational evaluation today is assessment based on student competencies. One of the innovative assessment methods is to add a type of objective multiple-choice questions in the student's ability assessment test. The outstanding advantage of the multiple-choice method is that it is easy to use many specialized software to analyze multiple-choice questions, test the task quickly and accurately such as determining the discriminant, difficulty, reliability, and relevance of the multiple-choice task with the Item Response Theory, as well as the factors regarding student competencies and their relationships can be described in the form of texts and graphs. The results of analyzing and evaluating the objective multiple-choice questions for assessing the physics competencies of students in teaching the Light Waves chapter of Physics grade 12 using the Item Response Theory are presented in detail in this report, with the support of Quest/Conquest software.

**Key words** - Item Response Theory; Quest/Conquest software; rated capacity; objective multiple-choice questions; physical competence

học tập,... và trong phương pháp viết người ta bổ sung các hình thức đánh giá khác như đưa thêm các dạng câu hỏi TN khách quan. Phương pháp kiểm tra TN khách quan đã được áp dụng nhanh chóng trong các kỳ kiểm tra, đánh giá quan trọng như: Kiểm tra giữa kỳ, kiểm tra cuối kỳ và kỳ thi tốt nghiệp trung học phổ thông (THPT), vì vậy đã thu hút sự quan tâm rất lớn từ các trường đại học sư phạm, các cơ sở quản lý giáo dục và đặc biệt là đội ngũ giáo viên. Trong các ưu điểm của phương pháp TN khách quan thì ưu điểm nổi bật là có thể sử dụng nhiều phần mềm chuyên dụng để phân tích các câu hỏi TN, đề TN một cách nhanh chóng và chính xác như xác định độ phân biệt, độ khó, độ tin cậy, độ giá trị cũng như mức độ phù hợp của đề kiểm tra với mô hình lý thuyết IRT. Nhóm tác giả nhận thấy, để nâng cao chất lượng dạy học, việc nghiên cứu và sử dụng phần mềm Quest/Conquest để phân tích và đánh giá các câu hỏi và đề TN là rất cần thiết và khoa học [1], [4], [8].

### 2. Kết quả nghiên cứu

#### 2.1. Lý thuyết ứng đáp câu hỏi IRT

Hiện nay, thông thường người ta phân loại lý thuyết TN thành hai dạng là lý thuyết khảo thí cổ điển và lý thuyết khảo thí hiện đại. Ngay khi vừa ra đời, lý thuyết khảo thí

<sup>1</sup> The University of Danang - University of Science and Education (Nguyen Bao Hoang Thanh)

<sup>2</sup> Chu Van An High School – Ninh Thuan (Nguyễn Thị Thảo)

cổ điển đã đạt được nhiều thành tựu, tạo nên cơ sở khoa học để thiết kế các phép đo tương đối chính xác. Tuy nhiên, lý thuyết này còn một số hạn chế như sau: Hạn chế cơ bản nhất của lý thuyết khảo thí cổ điển là không thể phân biệt được các đặc trưng của thí sinh độc lập với các đặc trưng của đề TN, đặc trưng này chỉ có thể giải thích trong mối quan hệ với đặc trưng kia. Hạn chế tiếp theo của lý thuyết khảo thí cổ điển nằm ở định nghĩa của độ tin cậy. Theo lý thuyết khảo thí cổ điển, độ tin cậy là “tương quan giữa các điểm của hai đề TN tương đương”. Trong thực tế không thể có các đề TN thỏa mãn tiêu chí tương đương. Một hạn chế khác của lý thuyết khảo thí cổ điển là nó tập trung vào khía cạnh đáp ứng ở mức độ đề TN chứ không phải ở mức độ câu hỏi TN. Do đó, thiếu cơ sở để xác định xem một thí sinh nào đó ứng đáp tốt ra sao đối với một câu hỏi đặt ra cho thí sinh đó [2].

Để khắc phục những hạn chế đó lý thuyết khảo thí hiện đại ra đời và dần thay thế cho lý thuyết khảo thí cổ điển. Lý thuyết khảo thí hiện đại với việc sử dụng lý thuyết Ứng đáp câu hỏi thường gọi là lý thuyết IRT (viết tắt của Item Response Theory) với mô hình Rasch là công cụ được ứng dụng nhiều nhất hiện nay. Trong bài viết này, nhóm tác giả tập trung vào các vấn đề liên quan đến lý thuyết khảo thí hiện đại. Lý thuyết IRT được xây dựng trên khoa học về xác suất và thống kê. Từ những năm 1970 trở đi, các công trình quan trọng trong lý thuyết IRT đã đạt được nhiều thành tựu, được công nhận và áp dụng phổ biến trong thực tiễn. Lý thuyết IRT được phát triển rất nhanh nhờ khả năng tính toán bằng máy vi tính và đạt được những thành tựu quan trọng nâng cao độ chính xác của các câu hỏi TN và đề TN. Lý thuyết IRT do nhà toán học Georg Rasch mô hình hóa mối quan hệ giữa mức độ năng lực của người làm TN và đáp ứng của người ấy với câu TN. Mỗi HS đứng trước một câu hỏi TN sẽ ứng đáp như thế nào, điều đó phụ thuộc vào năng lực tiềm ẩn của HS và các đặc trưng của câu hỏi. Hành vi ứng đáp này được mô tả bằng một hàm đặc trưng câu hỏi cho biết xác suất trả lời đúng câu hỏi tùy theo tương quan giữa năng lực của HS. Mỗi câu TN được mô tả bằng một thông số (độ khó) ký hiệu là  $\delta$  và mỗi người làm TN được mô tả bằng một thông số (năng lực) ký hiệu là  $\theta$ . 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 TN 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  $\theta$  trả lời đúng câu hỏi có độ khó  $\delta$  [3].

## 2.2. Sử dụng phương pháp TN khách quan đánh giá năng lực vật lý của HS

Trong các phương pháp được sử dụng để kiểm tra, đánh giá năng lực HS thì phương pháp kiểm tra viết là một phương pháp được sử dụng phổ biến cho cả hình thức đánh giá thường xuyên/đánh giá quá trình (đánh giá vì học tập; đánh giá là học tập) và đánh giá định kỳ/đánh giá tổng kết (đánh giá kết quả học tập). Và trong phương pháp viết thì phương pháp TN khách quan hiện nay đang được sử dụng rộng rãi ở các trường phổ thông trong các bài kiểm tra, đánh giá định kỳ, đặc biệt là ở khối lớp 12. Một bài kiểm tra TN

khách quan thường bao gồm nhiều câu hỏi, mỗi câu thường được trả lời bằng một dấu hiệu đơn giản hay một từ, một cụm từ. Câu hỏi TN khách quan bao gồm các loại sau [5]:

- Loại câu TN nhiều lựa chọn;
- Loại câu TN đúng – sai;
- Loại câu TN điền vào chỗ trống;
- Loại câu TN ghép đôi.

Hiện nay, ở các trường phổ thông loại câu hỏi TN nhiều lựa chọn được sử dụng phổ biến trong các bài kiểm tra, đánh giá thường xuyên và định kỳ và thì tốt nghiệp trung học phổ thông cho nhiều môn học. Trên cơ sở khoa học đánh giá, lý thuyết khảo thí hiện đại, nhóm tác giả đã xây dựng hệ thống câu hỏi TN khách quan nhiều lựa chọn chương “Sóng ánh sáng” – Vật lý 12 để đánh giá năng lực vật lý của HS trung học phổ thông. Nhóm tác giả đã tiến hành thử nghiệm các đề kiểm tra từ hệ thống câu hỏi TN khách quan đã xây dựng, sau đó phân tích kết quả thu được trên mô hình Rasch với phần mềm Quest/Conquest. Trong khuôn khổ bài báo này, sẽ trình bày kết quả phân tích thực nghiệm bài kiểm tra số 1 với 40 câu hỏi TN nhiều lựa chọn và thời gian làm bài 50 phút. Đề kiểm tra được thực nghiệm ở 380 HS khối 12 của trường THPT Chu Văn An, tỉnh Ninh Thuận.

Kết quả phân tích giúp lựa chọn những câu hỏi tốt, chỉnh sửa hoặc loại bỏ những câu hỏi chưa đạt yêu cầu để góp phần nâng cao chất lượng các câu hỏi và đề kiểm tra [6].

## 2.3. Xử lý dữ liệu bằng phần mềm Quest/Conquest

Quest và Conquest là hai phần mềm phân tích và đánh giá câu hỏi dựa trên lý thuyết IRT, nó cung cấp một phạm vi linh hoạt và toàn diện về các mô hình ứng đáp câu hỏi cho người phân tích, giúp khảo sát các thuộc tính về đánh giá năng lực của thí sinh làm bài kiểm tra. Phần mềm này có thể giúp chúng ta tính toán cho các kết quả liên quan đến mức độ phù hợp của các câu hỏi với mô hình Rasch, năng lực của thí sinh so với độ khó câu hỏi, độ tin cậy của đề thi và các chỉ số đặc trưng của 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, đồng thời cũng có thể chỉ ra kết quả bất thường của người làm câu TN nếu có [6], [7], [8].

## 2.4. Phân tích câu hỏi

### 2.4.1. Phân tích 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			.00
SD			.76
SD (adjusted)			.75
Reliability of estimate			.98
Fit Statistics			
=====			
Infit Mean Square		Outfit Mean Square	
Mean Square	1.00	Mean	1.00
SD	.05	SD	.07
Infit t		Outfit t	
Mean	.04	Mean	.07
SD	1.28	SD	.92
Summary of case Estimates			

```

=====
Mean                -.06
SD                  .77
SD(adjusted)       .68
Reliability of estimate .79
Fit Statistics
=====
Infit Mean Square   Outfit Mean Square
Mean               1.00           Mean               1.00
SD                 .13             SD                 .20
Infit t            Outfit t
Mean               -.01            Mean               .03
SD                 .99             SD                 .75
    
```

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

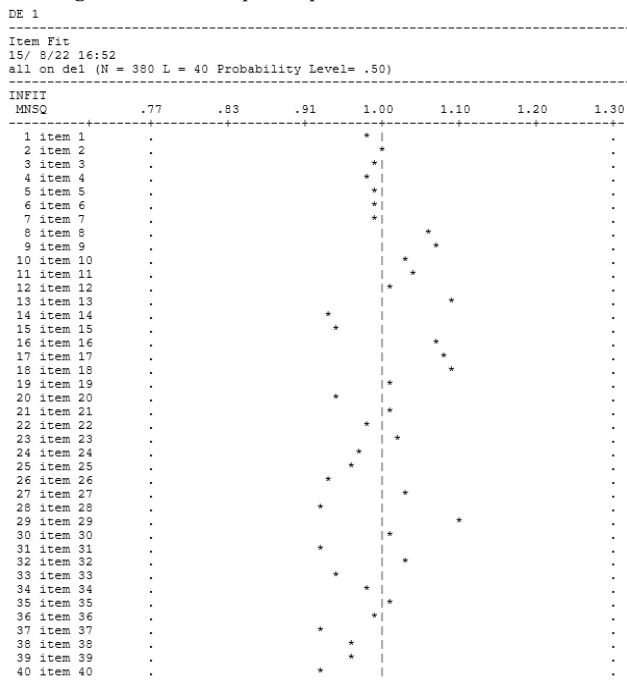
2.4.2. Độ tin cậy của đề kiểm tra

Kết quả tính toán bằng phần mềm Quest ở bảng Summary of item Estimates (Bảng 1) cho thấy, độ tin cậy của đề kiểm tra (Reliability of estimate) đạt 0,98. Qua đó chúng ta, đây là một kiểm tra có độ tin cậy cao và kết quả đo được phản ánh đúng năng lực của HS tham gia làm bài kiểm tra.

2.4.3. Mức độ phù hợp của các câu hỏi với mô hình Rasch

Trong biểu đồ Item Fit sau đây, mỗi câu TN biểu thị bằng dấu \*. Những câu TN 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 hỏi TN nào nằm ngoài khoảng này là không phù hợp và sẽ bị loại bỏ [6], [7], [8], [9].

Bảng 2. Minh họa sự phù hợp của các câu hỏi với mô hình



Kết quả cho thấy, 40 câu hỏi của đề kiểm tra nằm trong khoảng đồng bộ cho phép [6], [7], [8], [9].

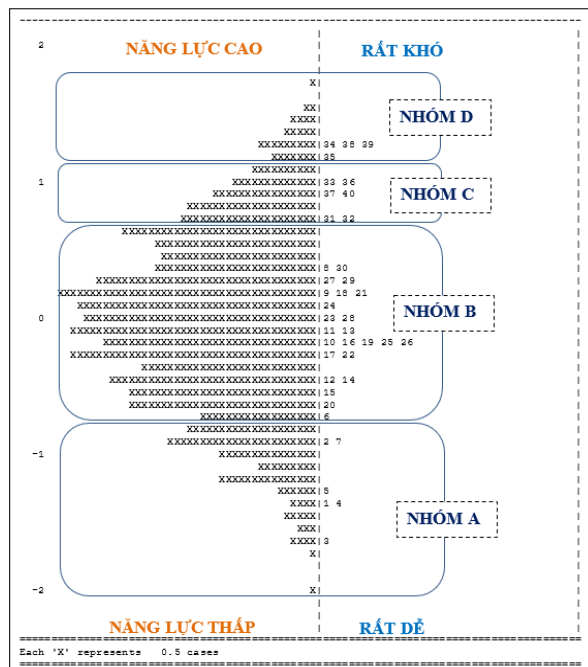
2.4.4. Phân bố năng lực HS và độ khó câu hỏi

Các thông tin về kết quả tính toán từ bảng Summary of case Estimates (Bảng 1) cho thấy năng lực trung bình của

mẫu thí sinh (case) tham gia bài kiểm tra (-0,06) nhỏ hơn và gần bằng so với độ khó chung của bài kiểm tra (0,00) cho thấy độ khó của đề thi tương đối phù hợp với năng lực của HS tham gia làm bài kiểm tra.

Sử dụng lý thuyết khảo thí hiện đại, năng lực của HS và độ khó của câu hỏi được đánh giá bằng thang đo logistic. Theo Bảng 3 các câu hỏi có độ khó nằm trong khoảng -1,6 đến 1,29. Trong khi đó, năng lực của thí sinh phân bố từ -1,72 đến 1,72 với trung bình cộng là -0,06 và độ lệch chuẩn 0,77. Điều này cho thấy, đề thi hơi khó so với năng lực của nhóm HS làm đề kiểm tra này.

Bảng 3. Phân bố năng lực HS và độ khó câu hỏi (thang logistic)



Kết quả xử lý bằng phần mềm Conquest cho sơ đồ phân bố thống kê tương quan giữa năng lực HS và độ khó câu hỏi trong đề kiểm tra như Bảng 3. Sơ đồ phân bố độ khó câu hỏi kiểm tra và năng lực HS cho thấy, mức độ phù hợp của đề kiểm tra đối với HS dự thi. Trong sơ đồ này các số bên phải đường chấm thẳng đứng trình bày sự phân bố các câu hỏi TN theo độ khó của từng câu trong bài TN. Những câu hỏi khó và những HS có năng lực cao được phân bố tiến dần lên phía trên (0,0), còn những câu hỏi dễ và những HS có năng lực thấp được phân bố tiến dần về phía dưới (0,0). Từ sơ đồ cho thấy, các câu hỏi khó chiếm 10% (Câu 34, 35, 38, 39); Câu hỏi ở mức tương đối khó chiếm 15% (Câu 31, 32, 33, 36, 37, 40); Câu hỏi mức độ trung bình chiếm 60% (Câu 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30); Và câu hỏi mức độ dễ chiếm 10% (Câu 1, 2, 3, 4, 5, 7). Do vậy, khả năng phân loại HS của đề kiểm tra này tương đối tốt. Bên cạnh đó, kết quả phân tích từ Hình 3 cho thấy đây là đề thi tương đối phù hợp với năng lực của HS tham gia làm bài kiểm tra, cụ thể:

- Nhóm HS có năng lực cao (Nhóm D) chiếm 3,95%, là những HS có khả năng trả lời đúng được những câu hỏi ở mức vận dụng (mức C) và mức vận dụng cao (mức D) theo thang đo Bloom. Ở mức này HS có thể trả lời đúng những

câu hỏi khó với mức ngưỡng từ 1,14. Cụ thể câu 38 có ngưỡng thresholds là 1,29:

**Câu 38: (N4, T3, V1.D)** Trong thí nghiệm Young về giao thoa ánh sáng, khoảng cách giữa hai khe là 2 mm, khoảng cách từ hai khe đến màn là 2 m. Chiều đến hai khe cùng lúc hai ánh sáng có bước sóng  $\lambda_1 = 0,4 \mu\text{m}$  và  $\lambda_2 = 600 \text{ nm}$ . Tại hai điểm M, N nằm khác phía trên màn quan sát, cách vân trung tâm lần lượt là 14,2 mm và 5,3 mm. Số vân sáng quan sát được trên đoạn MN của hai bức xạ là:

- A. 75      B. 72      C. 62      D. 65

Câu hỏi 38 dùng để đánh giá năng lực thành tố: Nhận thức kiến thức vật lý ứng với chỉ số biểu hiện hành vi (N4); Tìm tòi và khám phá thế giới tự nhiên dưới góc độ vật lý ứng với chỉ số biểu hiện hành vi (T3); Vận dụng kiến thức vật lý vào thực tiễn, chỉ số biểu hiện (V1) của năng lực vật lý.

- Nhóm HS có năng lực tương đối cao (Nhóm C) chiếm 11,28%. Ở mức năng lực này HS có thể trả lời đúng những câu hỏi tương đối khó, có ngưỡng thresholds từ 0,67 đến 1,01.

- Nhóm HS có năng lực trung bình (Nhóm B) chiếm 69,82%. Ở mức năng lực này HS có thể trả lời đúng những câu hỏi ở mức trung bình, có ngưỡng thresholds từ - 0,69 đến 0,38.

- Nhóm HS có năng lực thấp (Nhóm A) chiếm 14,95%, là những HS chỉ có khả năng trả lời đúng được những câu hỏi ở các mức độ nhận biết (mức A) theo thang đo Bloom. Ở mức này HS có thể trả lời những câu hỏi ở mức độ dễ, có ngưỡng thresholds từ - 1,6 đến - 0,89 [10]. Cụ thể, câu 3 có ngưỡng thresholds là -1,6:

### Câu 3: (N1.A)

Máy quang phổ là dụng cụ dùng để:

- A. Tổng hợp ánh sáng đa sắc từ các ánh sáng đơn sắc.  
B. Phân tích ánh sáng phức tạp thành các thành phần đơn sắc.  
C. Đo vận tốc của các ánh sáng đơn sắc.  
D. Đo bước sóng của các ánh sáng đơn sắc.

Câu hỏi 3 dùng để đánh giá năng lực thành tố: Nhận thức kiến thức vật lý ứng với hành vi biểu hiện bởi chỉ số (N1) của năng lực vật lý.

#### 2.4.5. Phân tích các tiêu chí khác

Ta tiếp tục xem xét các chỉ số thu được ở tập tin del.ita khi phân tích bằng phần mềm Quest như sau:

- *Categories*: Câu chọn TN, phương án đúng được đánh dấu (\*).

- *Missing*: Số HS không tìm được câu chọn, bỏ trống câu đó.

- *Disc*: Độ phân biệt của câu hỏi giữa các nhóm HS.

- *Percent*: Tỷ lệ phần trăm của một phương án là tỷ lệ giữa số HS chọn phương án đó so với tổng số HS làm bài kiểm tra.

- *Pt – Bserial*: 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, HS thực hiện được bước này khi trả lời đúng câu hỏi.

- *Thresholds*: Ngưỡng để vượt qua, thực chất là độ khó của câu TN.

- *P-value*: Độ tin cậy thống kê của độ phân biệt.

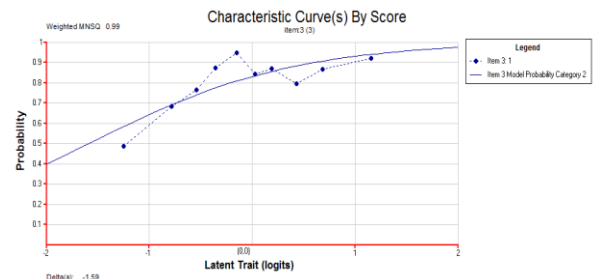
- *Mean ability*: Thang đo năng lực logarit của thí sinh đưa ra sự lựa chọn của mình.

- *Error*: Sai số trong tính toán.

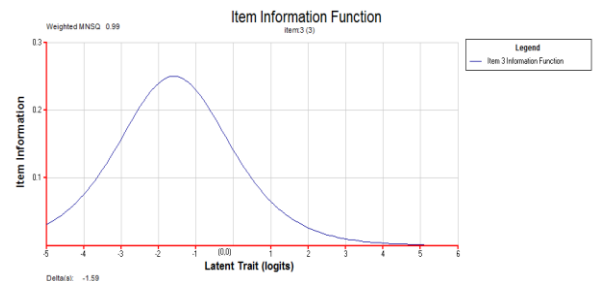
**Bảng 4.** Chỉ số thống kê của các câu hỏi thi được tạo ra từ Quest

Item	3: item 3				
					Infit MNSQ = .99 Disc = .26
Categories	A	B*	C	D	missing
Count	27	305	26	22	0
Percent (%)	7.1	80.3	6.8	5.8	
Pt-Biserial	-.15	.26	-.12	-.14	
p-value	.001	.000	.009	.002	
Mean Ability	-.49	.05	-.41	-.51	NA
Step Labels		1			
Thresholds		-1.60			
Error		.14			

Dựa vào Bảng 4 ta tiến hành phân tích các chỉ số của mỗi câu hỏi trong đề kiểm tra và đây là cơ sở khoa học để lựa chọn các câu hỏi TN có độ khó, độ phân biệt, độ tin cậy đạt yêu cầu để lưu vào ngân hàng câu hỏi dùng cho lần sau. Ví dụ, phân tích câu hỏi số 3 (item 3): Kết quả thu được ở Bảng 3 và 4 cho thấy, câu 3 là câu hỏi dễ nhất, có độ khó  $p = 0,803$  và độ khó  $\text{delta} = -1,59$  nên xác suất trả lời đúng của câu hỏi này rất cao thể hiện ở Hình 1, và đánh giá tốt nhất đối với nhóm HS có năng lực thấp (vùng năng lực từ -2,0 đến -1,0) thể hiện rất rõ ở Hình 2.



**Hình 1.** Đường cong đặc trưng câu hỏi số 3



**Hình 2.** Đường cong đặc trưng thông tin câu hỏi số 3

Ngoài ra, đây là câu hỏi có độ phân biệt chấp nhận được ( $D = 0,26$ ), câu hỏi có khả năng phân biệt được HS có năng lực cao và HS có năng lực thấp. Chỉ số Infit MNSQ = 0,99, nằm trong khoảng cho phép. Phương án A, C, D có độ phân biệt âm cho thấy các câu nhiễu tốt. Đối chiếu với ma trận của đề kiểm tra đã xây dựng thì đây là câu hỏi được thiết kế ở mức độ “Nhận biết” theo thang Bloom nên sẽ khá dễ, dùng đánh giá tốt nhóm HS có năng lực thấp. Vì vậy đây là câu hỏi TN đạt yêu cầu.

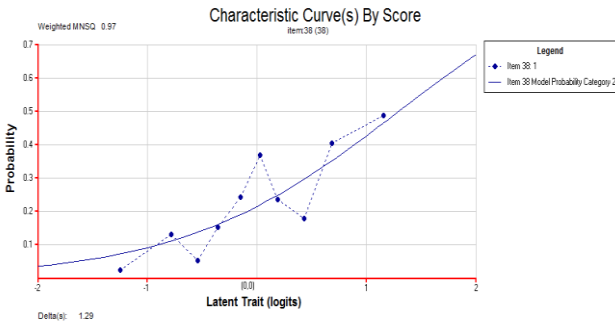
**\* Phân tích câu hỏi số 38 (item 38)**

**Bảng 5.** Chỉ số thống kê của câu hỏi 38 được tạo ra từ Quest

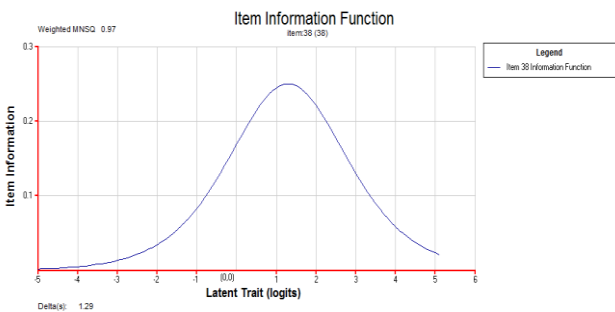
Item 38: item 38	Infit MNSQ = .96				
	Disc = .35				
Categories	A	B	C	D*	missing
Count	60	114	120	86	0
Percent (%)	15.8	30.0	31.6	22.6	
Pt-Biserial	-.11	-.03	-.20	.35	
p-value	.017	.307	.000	.000	
Mean Ability	-.25	-.09	-.28	.44	NA
Step Labels	1				
Thresholds	1.29				
Error	.13				

- Độ khó  $p = 0,226$ , đây là câu hỏi rất khó.
- Độ phân biệt  $D = 0,35$ : tốt.
- Infit MNSQ = 0,96, nằm trong khoảng cho phép.

Ngoài ra, phương án nhiều A, B, C có độ phân biệt âm cho thấy các câu nhiễu tốt. Qua đối chiếu với ma trận của đề kiểm tra này thì đây là câu hỏi thiết kế ở mức độ “Vận dụng cao” theo thang Bloom nên sẽ là câu hỏi khó, dùng đánh giá nhóm HS có năng lực cao. Và kết quả thu được ở Bảng 3 và 5 cho thấy, đây là câu hỏi khó nhất (có độ khó delta bằng 1,29) nên xác suất trả lời đúng câu hỏi này rất thấp, thể hiện qua Hình 3, và đánh giá tốt nhất nhóm HS có năng lực cao (vùng năng lực từ 1,0 đến 2,0) thể hiện rất rõ ở Hình 4. Như vậy, đây là một câu hỏi đạt yêu cầu.



**Hình 3.** Đường cong đặc trưng câu hỏi số 38



**Hình 4.** Đường cong đặc trưng thông tin câu hỏi số 38

**2.5. Phân tích năng lực HS**

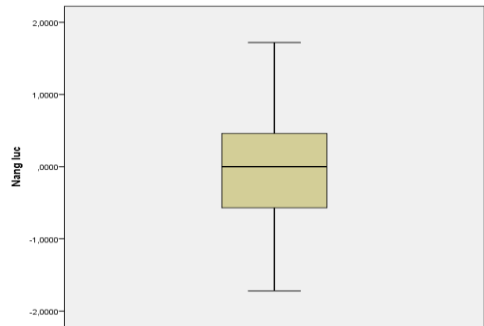
Để phân tích sâu hơn về năng lực của nhóm HS tham gia làm bài kiểm tra, nhóm tác giả sử dụng số liệu thu được sau khi phân tích bằng phần mềm Quest và đưa vào phần mềm chuyên dụng thống kê SPSS để tiến hành phân tích năng lực của nhóm HS làm bài kiểm tra. Kết quả phân tích năng lực của HS được trình bày ở Bảng 6:

**Bảng 6.** Bảng phân tích kết quả năng lực HS

Kết quả phân tích năng lực		
Biến năng lực		
Số lượng	Hợp lệ	380
	Bỏ sót	0
Trung bình		<b>-0,56342</b>
Trung vị		<b>,000000</b>
Số trội		<b>,2300</b>
Độ lệch chuẩn		,7682266
Phương sai		,590
Giá trị nhỏ nhất		<b>-1,7200</b>
Giá trị lớn nhất		<b>1,7200</b>
Phần trăm	25	<b>-,570000</b>
	50	<b>,000000</b>
	75	<b>,460000</b>

Từ Bảng 6 ta thấy, năng lực trung bình của nhóm HS tham gia làm bài kiểm tra này là -0,06, HS có năng lực thấp nhất là -1,72, HS năng lực cao nhất là 1,72, trung vị là 0 và năng lực của HS tập trung nhiều nhất ở mức 0,23; Có 25% trong tổng số HS có năng lực từ -0,75 trở xuống, 50% trong tổng số HS có năng lực từ 0 trở xuống và 75% trong tổng số HS có năng lực từ 0,46 trở xuống.

Dựa vào Hình 5 ta thấy, năng lực của nhóm HS phân bố khá đồng đều trong phạm vi một kỳ kiểm tra chuẩn lớp học. Năng lực trung bình của HS ở gần mức 0 [11].



**Hình 5.** Biểu đồ hình hộp – râu của biến năng lực

**2.6. Phân tích mối tương quan giữa năng lực HS và điểm thô**

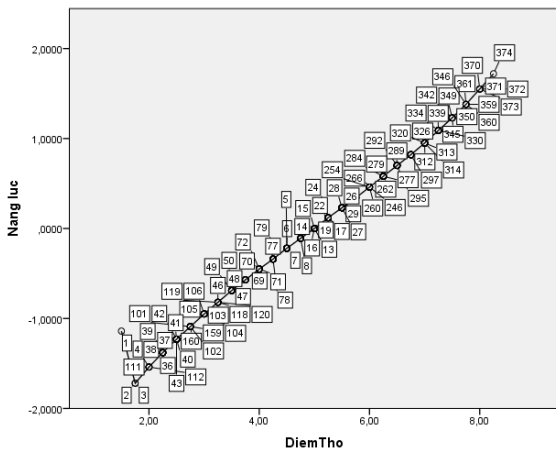
Sử dụng số liệu thu được sau khi phân tích bằng phần mềm Quest và điểm thô sau khi chấm bài làm của HS, nhóm tác giả sử dụng phần mềm SPSS để phân tích mối tương quan giữa năng lực HS và điểm thô, kết quả phân tích được trình bày ở Bảng 7.

**Bảng 7.** Bảng mối tương quan giữa năng lực và điểm thô

Mối tương quan			
		Năng lực	Điểm thô
Năng lực	Hệ số tương quan Pearson	1	<b>,999**</b>
	Mức ý nghĩa (2 chiều)		<b>,000</b>
	Số lượng	380	380
Điểm thô	Hệ số tương quan Pearson	<b>,999**</b>	1
	Mức ý nghĩa (2 chiều)	<b>,000</b>	
	Số lượng	380	380

\*\* . Mức ý nghĩa tại mức 0,01(2 chiều).

Qua kết quả thu được ở Bảng 7, ta thấy giữa hai đại lượng năng lực và điểm thô có sự tương quan rất cao (đến 0,999) với độ tin cậy của phép đo với mức ý nghĩa (hai chiều)=0. Do đó, có thể kết luận là điểm của kết quả bài kiểm tra thể hiện đúng năng lực thật sự của HS. Đồ thị Hình 6 càng thể hiện rõ hơn sự tương quan thuận này thông qua đường biểu diễn hướng đi lên.



Hình 6. Biểu đồ mối tương quan giữa năng lực và điểm thô

### 3. Kết luận

Các bảng số liệu trên minh họa kết quả phân tích 40 câu hỏi TN của một bài kiểm tra đánh giá năng lực vật lý có 380 HS tham gia, chứng tỏ rằng phần mềm Quest/Conquest rất tiện lợi để sử dụng, tiết kiệm thời gian và có hiệu quả trong việc phân tích số liệu các câu hỏi TN và đánh giá năng lực của HS theo lý thuyết khảo thí hiện đại IRT.

Nhờ sự hỗ trợ của hai phần mềm Quest/Conquest và SPSS, việc phân tích các câu hỏi TN đã trở nên nhanh chóng, tiện lợi và cung cấp kết quả đánh giá toàn diện cho đề TN. Kết quả phân tích thu được như sau: Chất lượng của các câu hỏi trong đề kiểm tra tương đối tốt, đề kiểm tra tương đối phù hợp với năng lực của nhóm HS tham gia làm bài kiểm tra.

Với phương pháp này kết hợp với các biểu đồ trên, chúng ta có thể thấy rằng đây là một công cụ dạy học hiện

đại không chỉ giúp phân tích và lựa chọn câu hỏi TN, để kiểm tra đạt yêu cầu lưu vào ngân hàng câu hỏi TN khách quan, mà còn giúp giáo viên đánh giá, phân tích được năng lực của HS tham gia làm bài kiểm tra. Đồng thời kết quả phân tích còn giúp giáo viên chẩn đoán và nắm bắt được tình hình học tập của các HS cá biệt, từ đó có thể giúp đỡ những HS này biết cách tự điều chỉnh quá trình học tập của mình.

### TÀI LIỆU THAM KHẢO

- [1] Nguyễn Bảo Hoàng Thanh, Sử dụng phối hợp trắc nghiệm khách quan và tự luận cải tiến hoạt động kiểm tra đánh giá kết quả học tập phần vật lý đại cương của sinh viên đại học sư phạm, *Luận án tiến sĩ giáo dục học*, Trường Đại học Vinh, 2002.
- [2] Lâm Quang Thiệp, *Đo lường trong giáo dục, lý thuyết và ứng dụng*, NXB Đại học Quốc gia Hà Nội, 2010.
- [3] Lê Thái Hưng, Nguyễn Văn Tuấn, Dương Thị Anh), “Xây dựng ngân hàng câu hỏi đánh giá kết quả học tập môn vật lý học kì I lớp 12: Vận dụng IRT và phần mềm Conquest”, *Tạp chí Quản lý Giáo dục*. Số 78 (tháng 12), 2015, p 40-45.
- [4] Nguyễn Bảo Hoàng Thanh, “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, p119-126.
- [5] Bộ GD & ĐT, *Tài liệu hướng dẫn bồi dưỡng giáo viên phổ thông cốt cán Mô đun 3 Kiểm tra đánh giá học sinh trung học phổ thông theo hướng phát triển phẩm chất, năng lực trong dạy học môn Vật lí*, Hà Nội, 2020.
- [6] Margaret L, Wu Raymond J, Adams Mark R, Wilson Samuel, *Handane. ACER conquest version 2.0: Generalised item response modelling software*, ACER press, 2007.
- [7] Phạm Xuân Thanh, *Bài giảng Mô hình Rasch và phân tích dữ liệu bằng phần mềm Quest*, Chương trình Đào tạo Thạc sĩ Đo lường và Đánh giá, Đại học Quốc gia Hà Nội 2013.
- [8] Nguyễn Thị ngọc Xuân, “Sử dụng phần mềm Quest/Conquest để phân tích câu hỏi trắc nghiệm khách quan”, *Diễn đàn trao đổi*. Số 12 (tháng 3), 2014, p24-27.
- [9] Adams, R. J., & Khoo, S. T, *Quest-Interactive test analysis system*, Victoria, Australia: Australian Council for Educational Research, 1996.
- [10] Wu M. L., Adams R. J., Wilson M. R, *ConQuest: Multi-aspect test software*. Camberwell: Australian Council for Education Research, 1997.
- [11] Fernandez, F. B. Action research in the physics classroom: the impact of authentic, inquiry based learning or instruction on the learning of thermal physics. *Asia-Pacific Science Education*, 3(1), 2017, 1–20.