

Sử dụng phần mềm quest/conquest để phân tích và nâng cao chất lượng đề kiểm tra tự luận dùng đánh giá năng lực của học sinh trung học phổ thông

Nguyễn Bảo Hoàng Thanh*, Võ Thị Hương**

*Trường Đại học Sư phạm – Đại học Đà Nẵng;

**Trường THPT Trần Phú - Đà Nẵng;

Received: 8/2/2023; Accepted: 14/2/2023; Published: 19/2/2023

Abstract: The 2018 education program is currently being implemented in grade 10, built according to the capacity development model. Quest/Conquest software is used for analyzing exams based on modern exam theories. It is grounded on Item Response Theory (IRT). This software provides information about the exam, the questions, students' ability and the relationship among those factors in the form of texts and graphs. In addition, it also gives quickly and accurately analysis results. This article presents those used results to analyse and evaluate the Physics 12 essay exam, ensure scientific and objective in the direction of assessing high school student's ability.

Keywords: Student; Quest; Conquest; Item Response Theory; ability.

1. Đặt vấn đề

Nếu công tác kiểm tra, đánh giá (KTĐG) kết quả học tập (KQHT) của học sinh (HS) được tổ chức và tiến hành một cách hợp lý, đúng đắn, đảm bảo tính khách quan, công bằng sẽ là động lực thúc đẩy học sinh chủ động, tích cực sáng tạo và không ngừng nâng cao chất lượng đào tạo. (CLĐT) [1], [2]

Hiện tại, ở các trường THPT, ở hầu hết tất cả các môn học, hình thức kiểm tra tự luận vẫn đang được sử dụng một cách thường xuyên và đều đặn. Bên cạnh đó, trong thực tế, việc phân tích, đánh giá đề tự luận vẫn còn nhiều hạn chế. Đề quá dễ dẫn đến kết quả bài làm của HS quá tốt nên bị xã hội phê phán là chạy theo thành tích, không phản ánh đúng năng lực của HS. Còn đề quá khó lại dẫn đến kết quả bài làm của HS quá kém nên bị xã hội phê phán là chất lượng giáo dục thấp. Tuy nhiên đánh giá CLGD mà chỉ nhìn vào điểm số là cách đánh giá rất thiếu chính xác [3], vì điểm số cao chưa chắc năng lực HS đó cao và điểm số thấp chưa chắc năng lực HS đó thấp. Vì vậy, vận dụng công cụ để đánh giá năng lực HS một cách có cơ sở khoa học là một yêu cầu cần thiết.

Ban đầu, phần mềm Quest ra đời, tuy xử lý dữ liệu nhanh hơn so với Excel nhưng chưa có nhiều hình ảnh đồ thị thông tin về đề kiểm tra, thông tin về câu hỏi và có giao diện console, là giao diện dòng lệnh, gây khó khăn trong việc thao tác. Từ đó, nhóm tác giả thuộc hai nền giáo dục lớn của thế giới là Úc và Mỹ đã khắc phục hạn chế này bằng việc phát triển

phần mềm Quest và cho ra đời phần mềm Conquest, với kết quả cho nhiều thông tin hơn.

Trong bài báo này, chúng tôi sẽ sử dụng lý thuyết khảo thí hiện đại, được xây dựng dựa trên lý thuyết ứng đáp câu hỏi (IRT) để phân tích, đánh giá một đề kiểm tra tự luận môn Vật lý 12 được thực nghiệm ở 160 HS của trường THPT Trần Phú, TP Đà Nẵng.

2. Cơ sở lý thuyết

2.1. Lý thuyết ứng đáp câu hỏi IRT

Lý thuyết ứng đáp IRT là lý thuyết được xây dựng trên cơ sở khoa học về xác suất và thống kê, dựa trên việc nghiên cứu mọi cặp tương tác nguyên tố: “Học sinh - câu hỏi (CH)”. Mỗi HS đứng trước một CH sẽ ứng đáp như thế nào phụ thuộc vào năng lực tiềm ẩn của HS và các đặc trưng của CH. Hành vi ứng đáp này được mô tả bằng hàm đặc trưng CH cho biết xác suất trả lời đúng CH tùy theo tương quan giữa năng lực HS và các tham số đặc trưng cho CH”. [4]

Trong lý thuyết ứng đáp IRT, ngoài mô hình khảo thí nhị phân (dichotomous), tương ứng với hai mức điểm 0 và 1, để đo lường chỉ một năng lực tiềm ẩn, hoặc nói cách khác, đo lường năng lực tiềm ẩn đơn chiều (unidimensional) còn có mô hình khảo thí đa phân đo lường năng lực tiềm ẩn đơn chiều, nhưng bằng các CH với kiểu ứng đáp đa phân (polytomous), tức là HS có thể ứng đáp không chỉ theo hai mức điểm 0 và 1, mà theo nhiều mức điểm khác nhau. Trong các mô hình khảo thí đa phân, mô hình định giá từng phần (partial credit model - PCM) được đặc

biệt chú ý vì tính tổng quát và khả năng ứng dụng rộng rãi của nó. Cụ thể là ứng dụng vào phân tích các đề kiểm tra tự luận được cho điểm từng phần.[5]

2.2. Xử lý dữ liệu bằng phần mềm Quest/Conquest

Đề kiểm tra này gồm 4 CH tự luận lớn được được tách ra thành 9 CH tự luận nhỏ. Sau khi kiểm tra xong, các bài làm của HS sẽ được chấm vào phiếu chấm chi tiết theo từng CH nhỏ theo đáp án đã được quy định, sau đó chuyển sang điểm nguyên với các mức gồm: 0, 1, 2, 3, 4.

24 mức điểm nguyên của 9 CH nhỏ cụ thể như sau: CH nhỏ 1, 2, 7 có các mức điểm nguyên: 0, 1, 2, 3, 4. CH nhỏ 4 có các mức điểm nguyên: 0, 1, 2, 3. CH nhỏ 5, 6, 8, 9 có các mức điểm nguyên: 0, 1, 2. CH nhỏ 3 có các mức điểm nguyên: 0, 1.

Dữ liệu được xử lý bằng phần mềm Quest và Conquest. Kết quả thu được gồm: độ tin cậy của CH, của đề kiểm tra, độ khó của CH, của đề kiểm tra, độ phân biệt, sai số các đề thị của CH và của đề kiểm tra, đồng thời phần mềm cũng chỉ ra kết quả bất thường của người làm CH nếu có.[6]

3. Kết quả nghiên cứu

3.1. Phân tích câu hỏi, đề kiểm tra tự luận

3.1.1. Phân tích mức độ phù hợp với mô hình

Summary of item Estimates

Mean	.00
SD	1.13
SD (adjusted)	1.07
Reliability of estimate	.89

Fit Statistics

Infit Mean Square	Outfit Mean Square
Mean 1.00	Mean .97
SD .10	SD .12

Summary of case Estimates

Mean	-.03
SD	1.37
SD (adjusted)	1.25
Reliability of estimate	.83

Từ file tuluan12.map (Quest) thu được kết quả tính toán ở bảng 3.1, cho thấy dữ liệu phù hợp với mô hình Rasch. Giá trị trung bình Mean bằng 1.00 và độ lệch chuẩn SD gần bằng 0.00. [8]

3.1.2. Độ tin cậy của đề kiểm tra

Từ file tuluan12.map (Quest), ở bảng Summary of item Estimates (bảng 1) cho thấy độ tin cậy của đề kiểm tra (Reliability of estimate) đạt 0.89. Qua

đó chứng tỏ đây là một đề kiểm tra có độ tin cậy cao và kết quả đo được phản ánh đúng năng lực của HS tham gia làm đề kiểm tra.[8]

3.1.3. Mức độ phù hợp của các câu hỏi

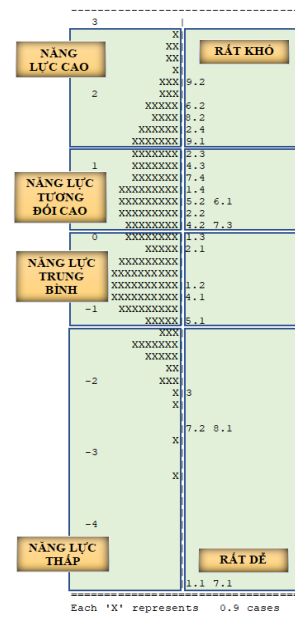
Từ file tuluan12.map (Quest), trong biểu đồ Item Fit ở hình 1, các CH được biểu thị bằng dấu *, ta thấy tất cả các CH đều nằm trong hai đường chấm thẳng đứng có giá trị INFIT MNSQ nằm trong khoảng (0.77; 1.30) cho thấy các CH phù hợp với mô hình Rasch, không có CH nào phải loại bỏ.[8]

INFIT MNSQ	.77	.83	.91	1.00	1.10	1.20	1.30
1 item 1	.		*				.
2 item 2	.		*				.
3 item 3	.			*			.
4 item 4	.					*	.
5 item 5	.		*				.
6 item 6	.			*			.
7 item 7	.				*		.
8 item 8	.				*		.
9 item 9	.					*	.

Hình 3.1. Mức độ phù hợp của các câu hỏi với mô hình Rash

3.1.4. Phân bố năng lực học sinh và độ khó câu hỏi

Từ file TL12.shw (Conquest) cho sơ đồ phân bố năng lực HS và độ khó CH kiểm tra như hình 3.2. Sơ đồ này cho biết mức độ phù hợp của đề kiểm tra đối với HS làm đề kiểm tra. Các số bên phải sơ đồ trình bày sự phân bố mức điểm. Còn các dấu x bên trái sơ đồ trình bày sự phân bố năng lực những HS tham gia làm đề.



Hình 3.2. Sơ đồ phân bố năng lực HS và độ khó câu hỏi (thang logistic)

Những mức điểm khó và những HS có năng lực cao được phân bố tiến dần lên phía trên (0.0), còn những mức điểm dễ và những HS có năng lực thấp được phân bố tiến dần về phía dưới (0.0).

Trong sơ đồ này cho biết cứ một dấu x đại diện cho có số lượng tương đương 0.9 HS tham gia làm bài. Kết quả cho thấy: có 5/24 mức điểm khó (chiếm 20.8%) gồm: 9.2, 6.2, 8.2, 2.4, 9.1; có 9/24 mức điểm tương đối khó (chiếm 37.6%) gồm: 2.3, 4.3, 7.4, 1.4, 5.2, 6.1, 2.2, 4.2, 7.3; có 5/24 mức điểm trung bình (chiếm 20.8%) gồm: 1.3, 2.1, 1.2, 4.1, 5.1; có 5/24 mức điểm dễ (chiếm 20.8%) gồm: 3, 7.2, 8.1, 1.1, 7.1.

Từ file tuluan12a.cas (Quest) và file TL12.itn (Conquest) thu được các kết quả sau: ngưỡng Thresholds trả lời đúng các câu hỏi là từ -32 đến 2.13 với giá trị trung bình là 0.00, trong khi mức năng lực của HS được phân bố từ -3.55 đến trên 3.17 với giá trị trung bình là -0.03 cho thấy độ khó của đề kiểm tra tương đối phù hợp với năng lực của HS tham gia làm đề kiểm tra. Cụ thể:

- Nhóm HS năng lực cao có mức năng lực từ 1.26 đến trên 3.17 (chiếm 21.9%), nhóm này trả lời trả lời được từ 18 đến 24 mức, trong đó có 5 mức điểm khó với ngưỡng Thresholds từ 1.39 đến 2.13.

- Nhóm HS năng lực tương đối cao có mức năng lực từ 0.3 đến 1.05 (chiếm 25%), nhóm này trả lời trả lời được từ 13 đến 17 mức, trong đó có 9 mức điểm tương đối khó có ngưỡng Thresholds từ 0.14 đến 1.13.

- Nhóm HS năng lực trung bình có mức năng lực từ -1.1 đến 0.11 (chiếm 35.6%), nhóm này trả lời trả lời được từ 7 đến 12 mức, trong đó có 5 mức điểm tương đối khó có ngưỡng Thresholds từ -1.22 đến 0.04.

- Nhóm HS năng lực thấp có mức năng lực từ -3.55 đến -1.48 (chiếm 17.5%), nhóm này trả lời trả lời được từ 3 đến 6 mức, trong đó có 5 mức điểm dễ có ngưỡng Thresholds từ -32 đến -2.2.

3.2. Phân tích năng lực học sinh

Sử dụng biểu đồ Kidmap từ file tuluan12.kid (Conquest), ngoài việc biết được năng lực của HS, nó còn cho phép ta nhanh chóng xác định được những HS nào có những câu trả lời bất thường. Cụ thể là HS nào làm đúng một số câu vượt quá năng lực (Harder Achieved), làm sai một số câu dưới mức năng lực (Easier not Achieved) hay làm đúng/sai trong mức năng lực của mình.[9]

Biểu đồ phân bố thống kê tương quan giữa độ khó và kết quả làm bài của HS có mã số 7 được trình bày ở bảng 3.3.

Từ bảng 3.3, ta thấy HS 7 là HS có mức năng lực thấp -3.55, có % score 4.55/100, có fit 1.39. Đây là HS có mức năng lực thấp nhất trong số 160 HS tham gia làm đề kiểm tra.

HS này trả lời đúng 3 mức trong đó có 1 mức điểm trả lời đúng ngoài dự kiến, 2 mức điểm trả lời đúng trong dự kiến.

Từ thực tế quan sát thấy rằng HS 7 thường xuyên vắng học, đi học trễ, trong giờ học ít phát biểu tham gia ý kiến tại lớp, làm bài tập về nhà không đầy đủ. Do đó, việc HS 7 có kết quả không cao là phản ánh đúng năng lực học vật lí của HS đó.

4. Kết luận

Từ những kết quả phân tích trên đây, ta có được một công cụ phân tích, đánh giá được chất lượng của đề kiểm tra tự luận như sau: Chất lượng đề kiểm tra tương đối tốt, phù hợp với mô hình, có độ tin cậy cao. Tuy nhiên, theo tác giả, đề vẫn còn 1 số câu có mức điểm khó so với năng lực HS, có thể xem lại và điều chỉnh CH nhỏ số 6, 8, 9 sao cho có thể đánh giá được HS ở mức năng lực trung bình.

KTĐG KQHT của HS là việc làm thường xuyên và liên tục. Nhờ có sự hỗ trợ của phần mềm Quest/Conquest, chúng ta đã phân tích các CH tự luận một cách nhanh chóng, chính xác, đánh giá được rất nhiều chỉ số về đề kiểm tra tự luận, về các CH, về năng lực của HS, về những câu trả lời bất thường của HS. Qua đó giúp người dạy có thêm thông tin ngược về chất lượng của các câu hỏi tự luận để hoàn thiện hơn các đề kiểm tra sử dụng trong quá trình dạy và học, góp phần nâng cao CLGD HS. Do đó, việc sử dụng phần mềm này cần được phổ biến rộng rãi và thực hiện thường xuyên bằng cách đưa vào các CTĐT ở các trường sư phạm và đưa vào các chương trình tập huấn hàng năm.

Tài liệu tham khảo

1. Bộ GD & Đ (2020,, *Tài liệu hướng dẫn bồi dưỡng giáo viên phổ thông cốt cán Mô đun 3 Kiểm tra đánh giá học sinh THPT theo hướng phát triển phẩm chất, năng lực trong dạy học môn Vật lí*, Hà Nội.

2. Nguyễn Bảo Hoàng Thanh, Nguyễn Thành Chiến (2021), *Bài giảng phân tích đề thi*, ĐHSP Phạm Đà Nẵng.

3. Phạm Xuân Thanh (2014), *Sử dụng hiệu quả các dạng thức câu hỏi thi - kiểm tra*. Bộ GD&ĐT.

4. Phạm Xuân Thanh (2013), *Bài giảng môn lý thuyết đo lường đánh giá cho học viên cao học chuyên ngành đo lường đánh giá*. Viện đảm bảo CLGD, ĐHQG Hà Nội.