

MỘT GIẢI PHÁP HỖ TRỢ PHÂN TÍCH, ĐÁNH GIÁ CÂU HỎI TRẮC NGHIỆM KHÁCH QUAN TRÊN HỆ THỐNG MOODLE THEO MÔ HÌNH IRT

Lê Thị Bích Hằng⁺,
Nguyễn Đình Hưng,
Phạm Thị Kim Ngoan,
Bùi Thị Hồng Minh

Trường Đại học Nha Trang
+ Tác giả liên hệ • Email: hangltb@ntu.edu.vn

Article history

Received: 05/6/2022

Accepted: 16/7/2022

Published: 20/8/2022

Keywords

Item Response Theory, IRT,
Item Characteristic Curve,
Moodle, MCQ Testing

ABSTRACT

Multiple-choice questions (MCQ) testing is a form of measurement and assessment that has been increasingly applied in schools. This form of assessment is being strongly supported on E-learning platforms, especially learning management systems. However, in Vietnam today, the quality assessment of MCQ items and tests in general, and online MCQ tests in particular, has not been given due attention. This paper introduces a solution to evaluate objective multiple-choice questions on Moodle systems using Item Response Theory. The solution is demonstrated through a Web application that the authors offer. Based on the analysis results, problematic questions requiring correction would be detected. This process can be iteratively conducted to improve the quality of the test items bank, thereby effectively improving the quality of the future MCQ tests.

1. Mở đầu

Trắc nghiệm khách quan (TNKQ) là một trong những hình thức đo lường và đánh giá trong giáo dục (Nguyễn Thị Ngọc Xuân, 2014). Các ưu điểm nổi bật của TNKQ bao gồm khả năng bao quát nội dung môn học cao và quá trình chấm thi khách quan, thuận tiện, nhanh chóng nhờ tự động hóa. Vì vậy, phương pháp đánh giá này đã và đang được áp dụng trong nhiều trường học trên thế giới để đo lường năng lực của người học, nhất là ở các kì thi có số lượng người dự thi quy mô lớn. Mặc dù TNKQ có nhiều ưu điểm, nhưng nếu ngân hàng câu hỏi không được biên soạn tốt và không được phân tích đánh giá, có khả năng một số câu hỏi kém chất lượng tiếp tục được sử dụng trong các bài thi sau, dẫn đến việc đánh giá sai lệch về năng lực của người học và có khả năng làm ảnh hưởng đến thành tích học tập của họ. Do đó, để đề thi có độ tin cậy và đo lường được năng lực của thí sinh, cần thiết phải có công cụ đánh giá câu hỏi, làm cơ sở nâng cao chất lượng các câu hỏi và đề thi. Với sự phát triển của công nghệ máy tính, lí thuyết ứng đáp câu hỏi (Item Response Theory - IRT) được ứng rộng ngày càng phổ biến và được xem là một trong những phương pháp hiệu quả trong việc phân tích câu hỏi và đề thi TNKQ cũng như đánh giá năng lực, kết quả học tập của thí sinh.

Hiện nay, trong kỉ nguyên Cách mạng công nghiệp 4.0, hầu hết các trường đại học đều triển khai mô hình học tập kết hợp (Blended learning) - là mô hình kết hợp giữa cách thức học truyền thống trên lớp và học trực tuyến - với sự hỗ trợ của một hệ thống quản lí học tập (Learning Management System - LMS); trong đó, Moodle là một trong những LMS được sử dụng phổ biến trên thế giới nói chung và Việt Nam nói riêng. Hệ thống Moodle cung cấp nhiều chức năng hỗ trợ giảng viên biên soạn câu hỏi và đề thi TNKQ trực tuyến; tuy nhiên, phương pháp phân tích câu hỏi và bài thi của Moodle chưa áp dụng IRT. Hơn nữa, mặc dù lí thuyết IRT có nhiều ưu điểm trong phân tích đánh giá đề thi, nhưng nó chưa được giảng viên ở các trường học trong nước biết đến rộng rãi. Ngoài ra, phần lớn giảng viên chưa chú trọng vào việc phân tích đánh giá câu hỏi trong đề thi TNKQ trực tuyến để hiệu chỉnh và hoàn thiện ngân hàng câu hỏi trên hệ thống LMS. Vì vậy, chúng tôi đã phát triển một ứng dụng Web làm công cụ hỗ trợ phân tích đánh giá câu hỏi trong bài thi TNKQ trực tuyến sử dụng mô hình IRT. Bước đầu chúng tôi phân tích các bài thi TNKQ sử dụng 02 dạng câu hỏi trắc nghiệm phổ biến là câu hỏi trắc nghiệm đa lựa chọn (multiple-choice question) có một phương án đúng và câu hỏi yêu cầu câu trả lời ngắn (short-answer question). Dựa trên kết quả phân tích, giảng viên có thể xác định những câu hỏi đạt yêu cầu nên giữ lại để sử dụng trong các đề thi sau và những câu hỏi chưa đạt yêu cầu cần được xem xét để hiệu chỉnh hoặc loại bỏ. Việc thực hiện quy trình này nhiều lần sẽ giúp loại

bỏ dần các câu hỏi có sai sót ra khỏi ngân hàng câu hỏi, cải thiện chất lượng câu hỏi và làm tăng hiệu quả của quá trình kiểm tra, đánh giá thông qua hình thức TNKQ trên hệ thống Moodle.

2. Kết quả nghiên cứu

2.1. Các phương pháp đánh giá câu hỏi trắc nghiệm khách quan

Trắc nghiệm là một phương pháp của khoa học về đo lường trong giáo dục. Một ưu điểm nổi bật của phương pháp trắc nghiệm là việc sử dụng các phần mềm để xử lý số liệu của câu hỏi trắc nghiệm. Có 02 cách tiếp cận chính để đánh giá câu hỏi TNKQ thông qua sử dụng dữ liệu phản hồi câu hỏi dựa trên các mẫu thử nghiệm (Lâm Quang Thiệp, 2010):

- Classical Test Theory - CTT: Sử dụng Lí thuyết trắc nghiệm cổ điển (Classical Test Theory - CTT) tập trung vào các tham số của câu hỏi là độ khó và độ phân biệt của câu hỏi, giá trị của các tham số được ước lượng phụ thuộc vào mẫu thí sinh được chọn để thử nghiệm và năng lực xác định được của thí sinh phụ thuộc vào đề kiểm tra.

- Item Response Theory - IRT: Sử dụng Lí thuyết ứng đáp câu hỏi (Item Response Theory - IRT) để ước lượng các tham số của đường cong đặc trưng câu hỏi (độ khó, độ phân biệt, độ đoán mò), các tham số này không phụ thuộc mẫu thử để định cỡ câu hỏi và năng lực đo được của thí sinh không phụ thuộc vào một đề kiểm tra cụ thể

Trong CTT, độ khó của câu hỏi là tỉ lệ phần trăm các thí sinh trả lời đúng câu hỏi đó, còn độ phân biệt mô tả mức độ ứng đáp đối với một câu hỏi của những nhóm thí sinh có năng lực cao và nhóm thí sinh có năng lực thấp. Với cách tính này, các tham số câu hỏi phụ thuộc vào năng lực của nhóm thí sinh, ví dụ: nếu bài kiểm tra được giao cho nhóm thí sinh ôn tập bài tốt hoặc có năng lực cao, thì bài kiểm tra được thực hiện rất dễ dàng; ngược lại, nếu bài kiểm tra được thực hiện bởi nhóm thí sinh không có kiến thức hoặc năng lực thấp, bài kiểm tra đó sẽ khó có điểm số cao. Đây là một trong những hạn chế của CTT bởi vì ước lượng các tham số câu hỏi có thể bị sai lệch do mẫu thí sinh được chọn.

Lí thuyết IRT xây dựng mô hình tính toán để xử lý dữ liệu dựa trên việc nghiên cứu các cặp tương tác giữa thí sinh và câu hỏi khi thực hiện một bài kiểm tra trắc nghiệm. Sự ứng đáp của mỗi thí sinh trước một câu hỏi phụ thuộc vào năng lực của thí sinh đó và các đặc trưng của câu hỏi. Kết quả phân tích IRT cho ra ba tham số ước lượng cho từng câu hỏi, lần lượt là a , b và c . Tham số a thể hiện độ phân biệt của câu hỏi, tham số b là độ khó của câu hỏi và tham số c để chỉ tỉ lệ đoán mò của thí sinh trả lời đúng câu hỏi. Ba tham số này là cơ sở cho việc chuẩn hóa và nâng cao chất lượng của từng câu hỏi nói riêng và đề thi nói chung. Từ ba tham số này, lí thuyết IRT được lượng hóa thành ba dạng mô hình như sau: (1) mô hình một tham số (mô hình Rasch) chỉ xét đến độ khó của câu hỏi; (2) mô hình hai tham số xét đến độ khó và độ phân biệt của câu hỏi; (3) mô hình ba tham số xét thêm mức độ đoán mò của thí sinh khi trả lời câu hỏi. Mỗi mô hình được thể hiện bằng một đường cong đặc trưng câu hỏi (Item Characteristic Curve - ICC) để xác định mối quan hệ giữa các đặc trưng của câu hỏi và năng lực của thí sinh và một hàm đặc trưng câu hỏi tương ứng để tính xác suất trả lời đúng câu hỏi ở một mức năng lực cụ thể (Baker, 2001).

Hàm đặc trưng câu hỏi (Item Characteristic Function - ICF) của các mô hình IRT có dạng sau:

- Mô hình một tham số (1PL):

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

- Mô hình hai tham số (2PL):

$$P(\theta) = \frac{1}{1 + e^{-a(\theta-b)}} \quad (2)$$

- Mô hình ba tham số (3PL):

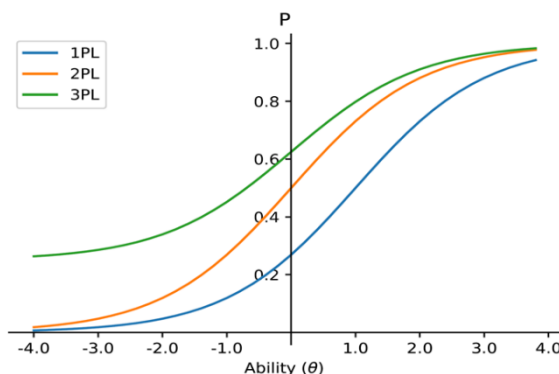
$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-a(\theta-b)}} \quad (3)$$

Trong đó: b là độ khó (difficulty); a là độ phân biệt (discrimination); c là độ đoán mò (guessing); θ là mức năng lực (ability); $P(\theta)$ là xác suất trả lời đúng câu hỏi ở một mức năng lực xác định.

Trong các mô hình trên, $P(\theta)$ là một hàm logistic; trong đó, các tham số của câu hỏi và năng lực của thí sinh đều chưa được biết trước. Do đó, cần thiết phải có một thuật toán phù hợp để ước lượng các tham số này. Các đường cong đặc trưng câu hỏi (ICC) của các mô hình IRT được minh họa ở hình 1; trong đó, các giá trị tham số của câu hỏi là $a = 1$, $b = 0$, và $c = 0,25$.

Hình dạng của ICC mô tả mối quan hệ giữa xác suất trả lời đúng một câu hỏi và thang đo năng lực thể hiện qua các tham số đặc trưng. Độ khó câu hỏi (b) là tham số xác định cách thức hoạt động của câu hỏi trên thang đo năng lực. Đây là một điểm trên thang đo năng lực mà xác suất trả lời đúng câu hỏi là 0,5 đối với mô hình một và hai tham số và là $(1 + c)/2$ đối với mô hình ba tham số. Trên đường cong ICC, các câu hỏi khó nằm lệch sang bên phải thang đo, thể hiện mức năng lực cao hơn của những thí sinh trả lời đúng câu hỏi, ngược lại những câu hỏi dễ nằm lệch sang

bên trái thang đo năng lực. Độ phân biệt của câu hỏi (a) được thể hiện thông qua độ dốc của đường cong ICC ở phần giữa của nó: đường cong càng dốc, câu hỏi càng có khả năng phân biệt tốt; đường cong càng phẳng, câu hỏi càng ít có khả năng phân biệt vì xác suất trả lời đúng ở mức năng lực thấp gần bằng với xác suất trả lời đúng ở mức năng lực cao. Trong mô hình ba tham số, giá trị của tham số đoán mò (c) được diễn giải trực tiếp vì nó là một xác suất, ví dụ: nếu $c=0,25$ có nghĩa là ở tất cả các mức năng lực, xác suất nhận được khi trả lời đúng câu hỏi bằng cách đoán mò là 0,25 (Baker, 2001).



Hình 1. Đường đặc trưng câu hỏi (ICC) của các mô hình IRT một tham số (1PL) với $b=0$; hai tham số (2PL) với $a=1, b=0$; và ba tham số (3PL) với $a=1, b=0$ và $c=0,25$

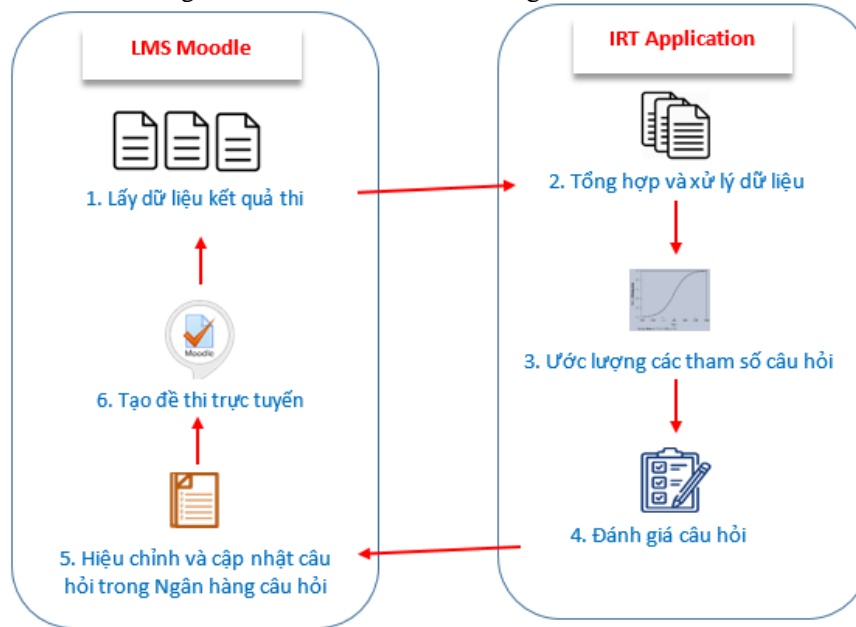
2.2. Ứng dụng web đánh giá câu hỏi trong bài thi trắc nghiệm khách quan

Hiện nay, để triển khai mô hình học tập hỗn hợp (Blended learning), nhiều trường đại học trong nước sử dụng hệ thống quản lý học tập Moodle để hỗ trợ giảng viên quản lý các khóa học của mình. Ngoài việc trang bị các chức năng cần có giúp giảng viên tạo câu hỏi và đề kiểm tra TNKQ trực tuyến, hệ thống Moodle tích hợp các tiện ích cung cấp số liệu phân tích thống kê kết quả bài làm của thí sinh. Tuy nhiên, các kết quả phân tích trong Moodle chủ yếu dựa vào CTT mà chưa áp dụng IRT (Butcher, 2010; Gamage et al., 2019). Hơn nữa, các kết quả thống kê này được thể hiện trên các giao diện riêng biệt và một đề kiểm tra được triển khai cho các lớp học khác nhau trên hệ thống Moodle sẽ nhận được các kết quả đánh giá khác nhau. Mặt khác, khi xây dựng đề kiểm tra TNKQ trên Moodle, hầu hết các giảng viên thực hiện theo quy trình: biên soạn câu hỏi đưa vào ngân hàng câu hỏi, tạo bài kiểm tra gồm các câu hỏi lấy từ ngân hàng có thiết lập chức năng xáo trộn câu hỏi và xáo trộn phương án, quy định điểm số mỗi câu hỏi và thời gian làm bài. Quy trình này chưa chú trọng vào việc phân tích đánh giá câu hỏi và đề thi để hiệu chỉnh và hoàn thiện ngân hàng câu hỏi trên hệ thống, do đó sẽ dẫn đến kết quả là những câu hỏi không đạt chất lượng vẫn còn được sử dụng trong các đề thi sau mà lẽ ra chúng phải được hiệu chỉnh hoặc loại bỏ. Hơn nữa, thực tế cho thấy các mô hình CTT và IRT chưa được các giảng viên trong trường biết đến rộng rãi, do đó việc phân tích đánh giá câu hỏi cũng như đề thi TNKQ và năng lực của thí sinh chưa được quan tâm.

Để hỗ trợ phân tích câu hỏi TNKQ, một số chương trình máy tính cung cấp ước lượng các tham số IRT cho nhiều môi trường khác nhau có thể kể đến là BILOG-3M, ConQuest, ConTEST, Winsteps. Mặc dù các chương trình này hỗ trợ mô hình trắc nghiệm nhị phân và đơn chiều, chúng có những hạn chế nhất định khi tích hợp dữ liệu kết quả thi lấy từ hệ thống Moodle vào các chương trình này để phân tích, vì các kết quả bài thi cần phải được xử lý dữ liệu thô một cách thủ công trước khi đưa vào phần mềm. Một giải pháp được đề xuất là xây dựng một trang Web sử dụng một công cụ phân tích mã nguồn mở nhận các file dữ liệu từ LMS, thực hiện quá trình ước lượng tham số câu hỏi bằng IRT và sau đó nhập lại kết quả phân tích vào LMS để hiển thị kết quả cuối cùng (Fotaris et al., 2010). Tuy nhiên, cách tiếp cận này ban đầu chỉ được triển khai thực nghiệm trên LMS Dokeos đồng thời có sự can thiệp vào mã nguồn và cấu trúc của LMS. Một giải pháp khác là thiết kế một ứng dụng Web sử dụng tập luật theo fuzzy logic để phân tích câu hỏi trong bài thi trắc nghiệm trực tuyến giúp giảng viên phát hiện các câu hỏi có vấn đề, tuy nhiên ứng dụng này không áp dụng mô hình IRT (Costagliola et al., 2014). Một đề xuất gần đây là xây dựng một ứng dụng phân tích đề thi trắc nghiệm trực tuyến theo cả hai mô hình CTT và IRT ba tham số (Mutawani et al., 2022), tập trung thử nghiệm đánh giá đề thi gồm các câu hỏi đa lựa chọn, mỗi câu hỏi có năm phương án, tuy nhiên ứng dụng này được thử nghiệm cho lớp học có số lượng tương đối ít và chưa triển khai cho các dạng câu hỏi trắc nghiệm khác.

Bài báo này giới thiệu một giải pháp phân tích, đánh giá câu hỏi trong đề thi TNKQ trên hệ thống Moodle theo mô hình IRT 2 tham số. Cụ thể, nhóm tác giả xây dựng một ứng dụng Web cho phép giảng viên dễ dàng nạp các file

dữ liệu thi được trích xuất từ Moodle, sau đó hệ thống tiến hành phân tích các câu hỏi và trả về kết quả thông qua các tham số ước lượng được. Kết quả phân tích giúp giảng viên có thêm thông tin khi cập nhật ngân hàng câu hỏi: giữ lại những câu hỏi đạt chất lượng, loại bỏ các câu hỏi có vấn đề hoặc sửa đổi chúng để sử dụng trong các đề thi sau. Giải pháp được đề xuất bao gồm các bước được thể hiện trong sơ đồ sau:



Hình 2. Quy trình đánh giá câu hỏi trong bài thi trắc nghiệm trực tuyến

- Bước 1: Từ hệ thống Moodle, các kết quả làm bài của thí sinh được xuất ra thành các tệp dữ liệu theo định dạng bảng tính Microsoft Excel.

- Bước 2: Các tệp dữ liệu này được tổng hợp và xử lý tự động trước khi đưa vào phân tích IRT. Nếu cùng một bài kiểm tra được thực hiện bởi các lớp học khác nhau thì ứng dụng cho phép bổ sung thêm các file dữ liệu của các lớp học khác.

- Bước 3: Ứng dụng thực hiện thuật toán Ước lượng biến cố hợp lý cực đại (Maximum likelihood estimation) (Mahmud, 2017) để tính toán các tham số a , b , và c đồng thời hiển thị đồ thị đường cong đặc trưng của mỗi câu hỏi.

- Bước 4: Dựa trên kết quả phân tích IRT, các câu hỏi được tô màu tùy theo độ khó của câu hỏi. Đối với những câu hỏi có độ phân biệt âm hoặc quá cao, chúng được gắn thêm nút cảnh báo.

- Bước 5: Các câu hỏi sau khi được hiệu chỉnh sẽ được cập nhật lại vào ngân hàng câu hỏi trên hệ thống Moodle. Bước này yêu cầu giảng viên hoặc người quản lý khóa học thực hiện.

- Bước 6: Sau khi nhóm câu hỏi trong bài kiểm tra bước đầu đã được hiệu chỉnh và đưa vào ngân hàng câu hỏi, giảng viên có thể sử dụng lại các câu hỏi này để đưa vào các đề kiểm tra sau.

Qua nhiều lần thực hiện quy trình trên, hầu như các câu hỏi bị lỗi sẽ bị loại bỏ khỏi ngân hàng câu hỏi hoặc được thay thế bằng những câu hỏi mới, nhờ vậy mà chất lượng của ngân hàng câu hỏi cũng như các đề kiểm tra sau này sẽ được cải thiện đáng kể.

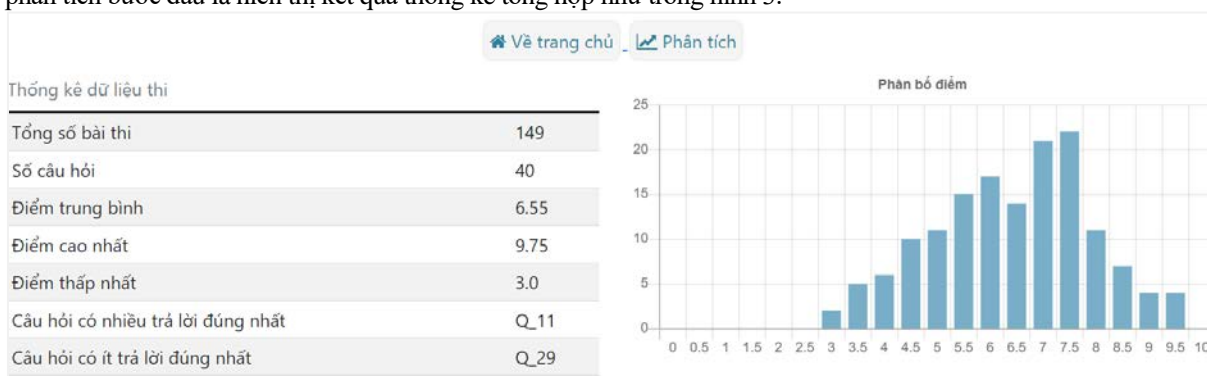
2.3. Minh họa kết quả thực nghiệm

Ứng dụng được sử dụng để phân tích dữ liệu các bài kiểm tra TNKQ trên hệ thống Moodle. Bước đầu, chúng tôi chọn mô hình IRT hai tham số để minh họa quá trình phân tích câu hỏi và các dạng câu hỏi trong các đề kiểm tra TNKQ có thể thuộc hai dạng: Câu hỏi đa lựa chọn (Multiple-choice question); Câu hỏi ngắn (Short-answer question). Dựa trên các kết quả phân tích, giảng viên có thể xem lại câu hỏi, hiệu chỉnh và cập nhật lại câu hỏi vào ngân hàng câu hỏi, từ đó có thể tạo các đề thi sau đạt chất lượng hơn.

Phạm vi lý thuyết của giá trị độ khó và độ phân biệt là từ $-\infty$ đến $+\infty$; nhưng về mặt thực tế thì giá trị của độ khó có phạm vi là $-3.0 \leq b \leq +3.0$ và giá trị của độ phân biệt là $-2.8 \leq a \leq +2.8$ (Baker, 2001). Một nhận xét khác là các câu hỏi được đánh giá tốt khi có độ phân biệt nằm trong phạm vi $0 < a < 2$; các câu hỏi có độ phân biệt nằm ngoài phạm vi này nên được xem xét hoặc loại bỏ, vì giá trị a âm hoặc lớn hơn 2 chỉ ra rằng các câu hỏi không thể phân

biệt chính xác những thí sinh có năng lực cao với những thí sinh có năng lực thấp (Hambleton, 1991). Từ các ngưỡng phân loại này, chúng tôi xếp loại câu hỏi theo ba mức độ: câu hỏi có độ khó thấp với $b \leq -1$, câu hỏi có độ khó trung bình với $-1 < b \leq 1$, câu hỏi có độ khó cao với $b > 1$. Với những câu hỏi có độ phân biệt $a \leq 0$ hoặc $a \geq 2$ được gắn thêm dấu hiệu cảnh báo ▲.

Để minh họa ứng dụng, tập dữ liệu được phân tích là kết quả điểm kiểm tra học phần Tin học đại cương A của sinh viên năm nhất trong học kỳ 1 năm học 2021-2022 tại Trường Đại học Nha Trang. Bài kiểm tra được triển khai trực tuyến trên 03 lớp học khác nhau với tổng số sinh viên tham gia làm bài là 149. Đề kiểm tra gồm 40 câu hỏi, trong đó có 37 câu hỏi dạng đa lựa chọn gồm 04 phương án trả lời và 03 câu hỏi dạng câu hỏi ngắn. Sau khi các file dữ liệu kết quả bài thi của sinh viên được kết xuất từ Moodle, chúng được đưa vào phần mềm để phân tích. Giai đoạn phân tích bước đầu là hiển thị kết quả thống kê tổng hợp như trong hình 3.

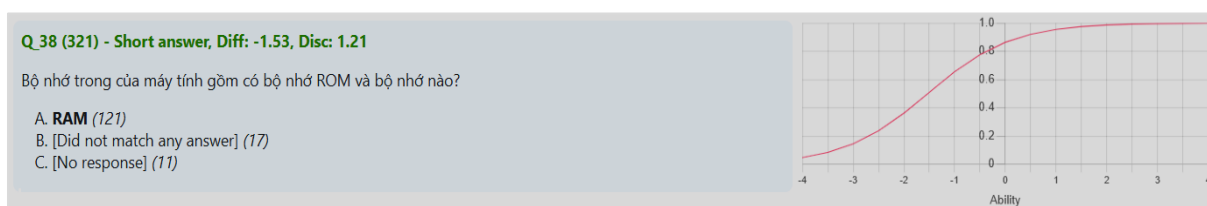


Hình 3. Giao diện thống kê tổng hợp kết quả bài thi

Khi nhấn nút lệnh *Phân tích*, ứng dụng tiến hành phân tích dữ liệu thi theo mô hình IRT hai tham số để ước lượng các tham số đặc trưng của mỗi câu hỏi. Kết quả phân tích chi tiết của các câu hỏi được hiển thị như trong hình 5.

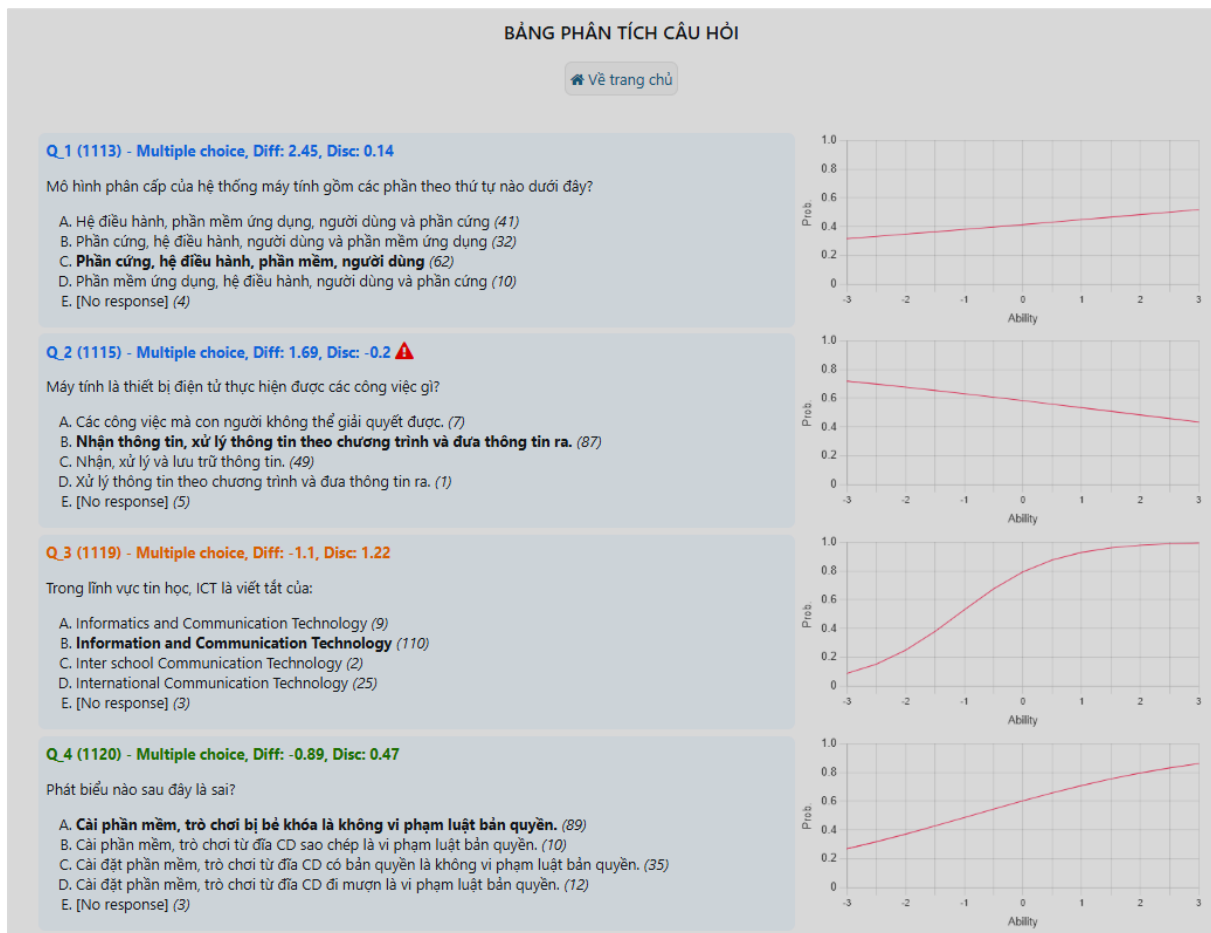
Các câu hỏi được phân loại theo màu như sau: câu hỏi có độ khó thấp, trung bình và cao được tô màu tương ứng lần lượt là cam, xanh lá cây và xanh dương. Ngoài ra, ứng dụng cung cấp thêm các thông tin của mỗi câu hỏi, bao gồm loại câu hỏi, giá trị độ khó (*Diff.*), giá trị độ phân biệt (*Disc.*) và nội dung câu hỏi. Đối với loại câu hỏi Multiple choice, sau nội dung mỗi phương án là số lượng thí sinh lựa chọn, phương án đúng được tô đậm, và “[No response]” là thông tin cho biết số thí sinh không chọn phương án nào để trả lời. Khi xét các phương án gây nhiễu, chúng chỉ nên được chọn bởi những thí sinh chưa nắm vững nội dung được đánh giá. Do đó, nếu một phương án nhiều được chọn bởi nhiều thí sinh, điều đó có thể cho thấy rằng câu hỏi hoặc các phương án lựa chọn không rõ ràng. Tương tự như vậy, nếu một phương án nhiều không có hoặc có rất ít thí sinh chọn, thì nó không đáng tin cậy và nên bị loại bỏ.

Đối với loại câu hỏi Short answer, ngoài thông tin đáp án được tô đậm và “[No response]” như câu hỏi dạng Multiple choice, thông tin “[Did not match any answer]” cho biết số lượng thí sinh đưa ra câu trả lời không khớp với đáp án (hình 4).



Hình 4. Minh họa thông tin chi tiết câu hỏi dạng trả lời ngắn

Đồ thị đường cong đặc trưng câu hỏi (ICC) thể hiện trực quan xác suất trả lời đúng câu hỏi dựa theo mức năng lực của thí sinh và các tham số của câu hỏi. Những câu hỏi có đồ thị hình chữ S và có độ dốc cho thấy câu hỏi đó phân biệt tốt các thí sinh, chẳng hạn như câu hỏi Q_3 (hình 5). Những câu hỏi có đồ thị dạng gần như tuyến tính và song song với trục năng lực cho thấy chúng có độ phân biệt thấp giữa các thí sinh có mức năng lực khác nhau, như câu hỏi Q_4 (hình 5). Hơn nữa, những câu hỏi có đường ICC giảm dần (như câu hỏi Q_2) biểu thị rằng thí sinh có điểm tổng thể cao hơn lại có khả năng trả lời đúng câu hỏi này thấp hơn thí sinh có điểm tổng thể thấp hơn và ngược lại. Vì thế, câu hỏi này có vấn đề, cần xem xét hiệu chỉnh hoặc loại bỏ khỏi ngân hàng đề thi.



Hình 5. Giao diện thể hiện thông tin chi tiết của một số câu hỏi sau khi phân tích theo mô hình IRT hai tham số

Đường cong đặc trưng câu hỏi cũng biểu thị trực quan độ khó của câu hỏi. Hình 5 cho thấy, câu hỏi Q_1 thuộc loại khó, vì thí sinh có mức năng lực rất cao (+3) chỉ có khả năng trả lời đúng là 0,5. Ngược lại, câu hỏi Q_3 thuộc loại dễ, vì thí sinh có mức năng lực thấp (-1) có xác suất trả lời đúng là 0,53. Ngoài ra, câu hỏi Q_4 có độ khó trung bình, vì thí sinh có mức năng lực trung bình (0) có xác suất trả lời đúng là 0,6. Như vậy, bảng kết quả phân tích câu hỏi như trên giúp giảng viên có cái nhìn trực quan về chất lượng các câu hỏi, từ đó làm cơ sở để đưa ra quyết định giữ lại câu hỏi đạt yêu cầu, chỉnh sửa hoặc loại bỏ câu hỏi có vấn đề ra khỏi ngân hàng câu hỏi.

3. Kết luận

Một bài kiểm tra TNKQ được xem là một thước đo để đánh giá hiệu quả giảng dạy cũng như năng lực của người học. Việc phân tích câu hỏi và bài kiểm tra có thể cung cấp thông tin có giá trị không chỉ về tính hợp lệ và độ tin cậy của bài kiểm tra mà còn về hiệu quả của quá trình học tập, giúp cải thiện mục tiêu học tập của người học. Chính vì vậy, với ngân hàng câu hỏi TNKQ đủ lớn được xây dựng trên một hệ thống LMS, các câu hỏi được đánh giá và thường xuyên được cập nhật sẽ có tác dụng tốt trong quá trình tạo các đề thi trắc nghiệm trực tuyến đạt chất lượng, làm giảm sự may rủi, “học tủ”, nâng cao trình độ và thái độ học tập tích cực của sinh viên.

Ứng dụng Web được đề xuất nhìn chung đã thực hiện một số chức năng cơ bản: tổng hợp các file dữ liệu về kết quả bài kiểm tra trên hệ thống Moodle, xử lý ước lượng các tham số của câu hỏi TNKQ dạng câu hỏi nhiều lựa chọn và câu hỏi trả lời ngắn, cung cấp thông tin hỗ trợ giảng viên hiệu chỉnh và cập nhật câu hỏi vào ngân hàng câu hỏi. Ngoài ra, ứng dụng cho phép tổng hợp dữ liệu kết quả thi từ các lớp khác nhau thực hiện cùng một bài thi trên hệ thống Moodle. Quy trình đánh giá đề thi TNKQ dựa trên phân tích câu hỏi như thế cho thấy ứng dụng này tương đối phù hợp khi được triển khai với các môn học được giảng dạy ở nhiều lớp, nhất là các môn học đại cương với số lượng sinh viên đông. Hơn nữa, ứng dụng có thể được sử dụng cho mọi học phần có áp dụng hình thức TNKQ với câu hỏi dạng multiple-choice, short-answer và true-false.

Tài liệu tham khảo

- Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation.
- Butcher, P. (2010). *Quiz report statistics*. Moodle. https://docs.moodle.org/dev/Quiz_report_statistics
- Costagliola, G., Ferrucci, F., & Fucella, V. (2014). *A Web-Based E-Testing System Supporting Test Quality Improvement*. Conference: Advances in Web Based Learning - ICWL 2007, 6th International Conference, Edinburgh, UK.
- Fotaris, P., Mastoras, T., Mavridis, I., & Manitsaris, A. (2010). Extending LMS to Support IRT-Based Assessment Test Calibration. *Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-642-13166-0_75
- Gamage, S. H. P. W., Ayres, J. R., Behrend, M. B., & Smith, E. J. (2019). Optimising Moodle quizzes for online assessments. *International Journal of STEM Education*, 6(1), 27. <https://doi.org/10.1186/s40594-019-0181-4>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications.
- Lâm Quang Thiệp (2010). *Đo lường trong giáo dục - Lý thuyết và ứng dụng*. NXB Đại học Quốc gia Hà Nội.
- Mahmud, J. (2017). Item response theory: A basic concept. *Educational Research and Reviews*, 12(5), 258-266.
- Mutiawani, V., Athaya, A. M., Saputra, K., & Subianto, M. (2022). Implementing Item Response Theory (IRT) Method in Quiz Assessment System. *TEM Journal*, 11(1), 210-218.
- Nguyễn Thị Ngọc Xuân (2014). Sử dụng phần mềm Quest/Conquest để phân tích câu hỏi trắc nghiệm khách quan. *Tạp chí Khoa học, Trường Đại học Trà Vinh*, 12, 24-27.