

VẬN DỤNG LÝ THUYẾT ỨNG ĐÁP CÂU HỎI PHÂN TÍCH DỮ LIỆU ĐỀ THI TRẮC NGHIỆM KHÁCH QUAN

Lê Thái Hưng
Trường Đại học Giáo dục - Đại học Quốc gia Hà Nội
Bùi Đức Quý, Nguyễn Thị Thoa
Học viện Công nghệ Bưu chính Viễn thông
Nguyễn Xuân Quy
Trường Đại học CMC
Lê Kiều Hưng
Trường Tiểu học, THCS, THPT Vinschool

Tóm tắt: Trong bối cảnh giáo dục hiện đại đòi hỏi yêu cầu cao về công tác kiểm tra đánh giá trong giảng dạy, việc áp dụng các phần mềm phân tích dữ liệu nhằm đánh giá chất lượng đề thi và năng lực thí sinh ngày càng trở nên cần thiết. Nghiên cứu này tập trung vào việc sử dụng phần mềm ConQuest để phân tích dữ liệu kết quả bài thi trắc nghiệm khách quan theo mô hình Rasch. Thông qua việc xử lý dữ liệu điểm số từ các bài thi thực tế, bài báo trình bày quy trình nhập liệu, chạy mô hình, và diễn giải các thông số đầu ra của phần mềm như độ phù hợp mô hình (fit statistics), độ khó, độ phân biệt của các câu hỏi, và năng lực của thí sinh. Kết quả nghiên cứu cho thấy ConQuest là công cụ hữu hiệu trong việc hỗ trợ phân tích, xây dựng và hiệu chỉnh đề thi trắc nghiệm một cách khoa học, khách quan và có cơ sở thống kê vững chắc. Nghiên cứu đồng thời đề xuất một số khuyến nghị nhằm nâng cao hiệu quả ứng dụng phần mềm này trong thực tiễn giảng dạy và kiểm tra đánh giá.

Từ khóa: Trắc nghiệm khách quan, Conquest, phân tích dữ liệu đề thi

DATA ANALYSIS OF OBJECTIVE MULTIPLE-CHOICE EXAM RESULTS USING CONQUEST SOFTWARE

Le Thai Hung
University of Education, Vietnam National University, Hanoi
Bui Duc Quy, Nguyen Thi Thoa
Posts and Telecommunications Institute of Technology
Nguyen Xuan Quy
CMC University
Le Kieu Hung
Vinschool Primary, Middle, and High School

Abstract: In the context of modern education, where high standards are required for assessment and evaluation in teaching, the use of data analysis software to evaluate test quality and candidate ability has become increasingly essential. This study focuses on employing ConQuest software to analyze data from objective multiple-choice test results using the Rasch model. By processing score data from actual tests, the paper outlines the procedures for data input, model execution, and interpretation of output parameters, including model fit statistics, item difficulty, item discrimination, and candidate ability. The findings indicate that ConQuest is an effective tool for supporting the scientific, objective, and statistically robust analysis, construction, and refinement of multiple-choice tests. The study also proposes recommendations to enhance the practical application of this software in teaching and assessment practices.

Keywords: Objective testing, Conquest, test data analysis

Nhận bài: 16/05/2025

Phản biện: 07/06/2025

Duyệt đăng: 12/06/2025

I. ĐẶT VẤN ĐỀ

Môn học Pháp luật đại cương là môn học bắt buộc đối với sinh viên trong các trường đại học (Bộ GD-ĐT, 1995), Học phần Pháp luật đại cương là học phần mà các tri thức, chuẩn mực, kỹ năng đều gắn với các sự kiện thực tế, các tình tiết có thực đang diễn ra. Giảng dạy học phần Pháp luật đại cương là một quá trình mà GV phải tổ chức, hướng dẫn SV hoạt động tự chiếm lĩnh các giá trị, chuẩn mực đạo đức, pháp luật thông qua việc nắm tri thức, qua thực hành rèn luyện trong và ngoài giờ học. Thông qua đó giúp SV có những phương thức

ứng xử về đạo đức, pháp luật, văn hóa, hướng dẫn cho SV thực hành trong cuộc sống hàng ngày (Nguyễn Thị Khuyến, 2017). Pháp luật đại cương là học phần bắt buộc thuộc khối kiến thức giáo dục đại cương trong chương trình đào tạo sinh viên đại học. Học phần cũng giúp người học tìm hiểu, nghiên cứu kiến thức cơ bản thuộc các ngành luật trong hệ thống pháp luật Việt Nam như chế độ chính trị, kinh tế, văn hóa, quyền con người, quyền và nghĩa vụ của công dân, bộ máy nhà nước Việt Nam, quyền sở hữu, quyền thừa kế, hợp đồng, vi

phạm pháp luật hành chính, tội phạm, và quy định pháp luật về phòng chống tham nhũng. Nghiên cứu này áp dụng lý thuyết khảo thí hiện đại (IRT) để tiến hành phân tích dữ liệu kết quả bài thi trắc nghiệm khách quan gồm 40 câu hỏi của học phần Pháp luật đại cương của sinh viên một trường đại học tại Hà Nội (không tiết lộ thông tin về Trường theo yêu cầu), qua đó nhận xét chi tiết công tác xây dựng đề thi và sự quan tâm của Nhà trường đối với học phần này.

II. CƠ SỞ LÝ LUẬN

2.1 Lý thuyết Khảo thí

Trong lĩnh vực đo lường và đánh giá giáo dục, hai hệ thống lý thuyết nền tảng thường được sử dụng để xây dựng và phân tích công cụ đánh giá là Lý thuyết kiểm tra cổ điển (Classical Test Theory - CTT) và Lý thuyết ứng đáp câu hỏi (Item Response Theory - IRT). Mỗi lý thuyết mang những quan điểm khác nhau về bản chất của điểm số, phương pháp đánh giá và cách mô hình hóa mối quan hệ giữa thí sinh và câu hỏi. Lý thuyết kiểm tra cổ điển, hình thành từ đầu thế kỷ XX và phát triển mạnh trong suốt nhiều thập kỷ sau đó, giả định rằng mỗi điểm số quan sát được (Observed Score - X) là tổng của năng lực thật sự của người học (True Score - T) và sai số ngẫu nhiên (Error - E), theo công thức: $X = T + E$ (Lord & Novick, 1968). CTT đóng vai trò nền tảng cho nhiều công cụ đánh giá truyền thống, góp phần quan trọng trong việc chuẩn hóa đề thi và phân tích độ tin cậy (reliability) cũng như độ phân biệt (discrimination) của các mục hỏi. Tuy nhiên, một hạn chế cơ bản của CTT là sự phụ thuộc lẫn nhau giữa đặc điểm của đề thi và năng lực thí sinh – điều này gây khó khăn khi so sánh giữa các nhóm khác nhau hoặc khi tái sử dụng câu hỏi trong các ngữ cảnh khác nhau (Crocker & Algina, 2008). Để khắc phục các nhược điểm này, từ những năm 1970, Lý thuyết ứng đáp câu hỏi (IRT) đã nổi lên như một mô hình phân tích hiện đại và mạnh mẽ hơn. Lý thuyết này, xuất phát từ công trình của nhà toán học Georg Rasch, xem khả năng trả lời đúng một câu hỏi là hàm số của năng lực cá nhân và đặc điểm của từng mục hỏi như độ khó (difficulty), khả năng phân biệt (discrimination), và xác suất đoán đúng (guessing) (Embretson & Reise, 2000). Trong mô hình IRT một tham số (1PL - Rasch model), xác suất thí sinh trả lời đúng phụ thuộc vào chênh lệch giữa năng lực của thí sinh và độ khó của câu hỏi. Với các mô hình hai và ba tham số (2PL, 3PL), hàm xác suất còn được điều chỉnh theo hệ số phân biệt và yếu tố đoán mò.

Nhờ sự hỗ trợ của công nghệ máy tính, IRT đã trở thành nền tảng cho các hệ thống thi trên máy tính (Computerized Adaptive Testing - CAT), nâng cao độ chính xác và hiệu quả trong đánh giá năng lực cá nhân (Van der Linden & Glas, 2000).

2.2 Bài kiểm tra trắc nghiệm khách quan (TNKQ)

Trắc nghiệm khách quan (TNKQ) là một trong những hình thức đo lường và đánh giá quan trọng trong giáo dục (Nguyễn Thị Ngọc Xuân, 2014). Bài kiểm tra trắc nghiệm khách quan là một phép thử (kiểm tra) để nhận dạng, xác định thu nhận những thông tin phản hồi về những khả năng, đặc tính, tính chất của một sự vật hay hiện tượng nào đó. Bài kiểm tra trắc nghiệm trong giảng dạy là một phép thử nhằm đánh giá khách quan trình độ, năng lực cũng như kết quả học tập của người học trước, trong quá trình và khi kết thúc một giai đoạn học tập nhất định (một phần hoặc bài giảng lý thuyết hoặc thực hành); một chương hoặc một chương trình đào tạo (Sái Công Hồng cùng cộng sự, 2017). Bắt đầu từ năm 2006, Bộ Giáo dục và Đào tạo tổ chức thi TNKQ cho môn Ngoại ngữ và từ năm 2007 tăng thêm các môn Vật lý, Hóa học và Sinh học trong các kỳ thi tốt nghiệp trung học phổ thông và đại học (Bùi Ngọc Quang, 2017). Hình thức kiểm tra đánh giá bằng đề thi trắc nghiệm khách quan trở nên phổ biến ở thời điểm hiện tại trong nhiều kỳ thi nhằm đánh giá năng lực, mức độ nắm bắt kiến thức của người học. Ngân hàng câu hỏi trắc nghiệm khách quan (CHTN) được xem là một trong những công cụ cốt lõi trong lĩnh vực kiểm tra đánh giá hiện đại. Theo McCallon và Schumacker (2002), ngân hàng câu hỏi là tập hợp các mục hỏi dưới dạng văn bản, được lưu trữ và truy cập dễ dàng, thường được số hóa nhằm hỗ trợ thuận lợi trong việc thiết kế các bài kiểm tra. Một hệ thống ngân hàng câu hỏi được thiết kế khoa học sẽ hỗ trợ hiệu quả cho giáo viên trong việc xây dựng đề kiểm tra phù hợp với các kỳ đánh giá khác nhau. Đồng thời, nó giúp giảm thiểu khối lượng công việc thủ công, tăng cường độ chính xác, nhất quán và tính hệ thống trong công tác kiểm tra, đánh giá kết quả học tập của học sinh (Lê Thái Hưng cùng cộng sự, 2015).

2.3 Phần mềm Conquest

Phần mềm ConQuest do Hội đồng nghiên cứu giáo dục Úc (ACER) thiết kế, ứng dụng lý thuyết ứng đáp câu hỏi IRT (Item Response Theory) vào phân tích dữ liệu đề thi và đánh giá năng lực của thí sinh (Acer Press, 2007). Để sử dụng phần mềm, cần có file dữ liệu kết quả làm bài thi trắc

nghiệm của thí sinh dạng đuôi file *.dat và file chạy lệnh phân tích *.cqc. Kết quả chạy phân tích dữ liệu đề thi gồm các file có đuôi *.itn, *.shw, *.mle, *.case. Lý thuyết Ứng đáp Câu hỏi (Item Response Theory - IRT) là một mô hình thống kê hiện đại được sử dụng để phân tích dữ liệu đánh giá, tập trung vào mối quan hệ giữa năng lực của thí sinh và xác suất trả lời đúng một câu hỏi cụ thể (Hambleton & Swaminathan, 1985). Khác với lý thuyết khảo thí cổ điển (Classical Test Theory - CTT), IRT cho phép mô hình hóa đặc điểm của từng câu hỏi (độ khó, độ phân biệt) và năng lực của thí sinh trên một thang đo chung, đảm bảo tính bất biến của các tham số ước lượng bất kể đặc điểm của nhóm thí sinh hay bài kiểm tra (Embretson & Reise, 2000). Trong phần mềm ConQuest, các mô hình IRT phổ biến như mô hình 1 tham số (Rasch), 2 tham số (2PL) được tích hợp để phân tích dữ liệu, cho phép đánh giá chính xác hơn về độ tin cậy và tính giá trị của bài thi. Quá trình phân tích trên ConQuest bắt đầu bằng việc nhập dữ liệu từ file *.dat, chứa kết quả bài thi trắc nghiệm của thí sinh dưới dạng mã hóa, và file lệnh *.cqc, trong đó quy định các thông số phân tích như mô hình IRT được chọn, số lượng tham số, và các yêu cầu báo cáo. Sau khi chạy phân tích, ConQuest tạo ra các file đầu ra chính bao gồm: file *.itn cung cấp thông tin chi tiết về các tham số của câu hỏi, như độ khó, độ phân biệt, file *.shw trình bày kết quả tổng quan về các tham số của bài thi và năng lực thí sinh, hỗ trợ việc đánh giá tính phù hợp của mô hình, file *.mle chứa các giá trị ước lượng năng lực của thí sinh theo phương pháp cực đại hóa khả năng (Maximum Likelihood Estimation), file *.case cung cấp thông tin chi tiết về từng thí sinh, bao gồm điểm số và các giá trị năng lực ước lượng. Các kết quả này đánh giá năng lực của từng thí sinh, hỗ trợ phân tích chất lượng của bài thi, xác định các câu hỏi có độ khó không phù hợp hoặc độ phân biệt thấp, từ đó đề xuất cải tiến trong thiết kế đề thi (Wright & Stone, 1979).

III. NỘI DUNG NGHIÊN CỨU

3.1. Phương pháp nghiên cứu

Nghiên cứu áp dụng lý thuyết ứng đáp câu hỏi (IRT) thông qua việc sử dụng phần mềm Conquest để phân tích dữ liệu kết quả bài thi trắc nghiệm khách quan môn Pháp luật đại cương, gồm dữ liệu đề thi của 301 thí sinh tại một trường đại học ở Hà Nội (thông tin cơ sở giáo dục không được tiết lộ theo yêu cầu). Đề thi gồm 40 câu hỏi trắc nghiệm khách quan, mỗi câu có 4 lựa chọn (A, B, C, D) với duy nhất 1 phương án đúng. Không áp dụng

điểm âm cho phương án sai. Tổng thời gian làm bài là 60 phút. Đề được xây dựng trên cơ sở kết hợp cả hai hướng: phân tầng theo mức độ nhận thức và phân bố theo chủ đề nội dung giảng dạy. Cụ thể, 30% số câu hỏi thuộc mức độ nhận biết, 35% thuộc mức thông hiểu và 35% còn lại là mức vận dụng, đảm bảo đánh giá đa dạng các cấp độ tư duy của sinh viên. Về mặt nội dung, đề thi bao phủ 4 chủ đề chính của học phần Pháp luật đại cương: (1) Tổng quan về Nhà nước và pháp luật; (2) Hệ thống các ngành luật cơ bản; (3) Quyền và nghĩa vụ pháp lý của công dân; và (4) Ứng dụng kiến thức pháp luật vào thực tiễn.

3.2. Kết quả nghiên cứu

Sau khi chạy phân tích dữ liệu kết quả đề thi bằng phần mềm Conquest, nhận được kết quả cụ thể như sau: chỉ số độ tin cậy phân tách (Separation Reliability) = 0.980, chứng tỏ đề thi có khả năng phân biệt tốt mức độ năng lực của thí sinh, ít bị nhiễu, ít đoán mò hoặc lỗi ngẫu nhiên ảnh hưởng đến kết quả, các tham số ước lượng (năng lực của người học) đáng tin cậy, có thể sử dụng cho đánh giá, phân loại, xếp hạng. Dữ liệu có phân phối gần chuẩn (các giá trị Skewness và Kurtosis xấp xỉ giá trị 0). Tuy nhiên, độ tin cậy của thang đo rất thấp (Cronbach's Alpha = 0.28), chứng tỏ mức độ thống nhất nội tại của bộ đề thi có vấn đề, cần rà soát lại các câu hỏi và các mục trong thang đo hoặc xây dựng lại cấu trúc đo lường.

Có 309 thí sinh tham gia kỳ thi, điểm trung bình là 17,77 điểm theo thang điểm 40 (mỗi câu hỏi đúng được 1 điểm) (tương đương với khoảng 4,4 điểm theo thang điểm 10) - một điểm số rất thấp, độ lệch chuẩn là 3.22, điều này cho thấy điểm số không quá phân tán, phần lớn tập trung quanh mức 17.77. Độ lệch (Skewness) có giá trị 0.24 cho thấy phân phối gần như đối xứng, chỉ hơi lệch phải nhẹ, nghĩa là có nhiều học sinh có điểm thấp hơn một chút so với trung bình. Độ nhọn (Kurtosis) Giá trị 0.04 gần bằng 0, cho thấy phân phối gần như chuẩn, không có quá nhiều điểm cực cao hoặc cực thấp. Độ tin cậy nội tại (Coefficient Alpha) của bài kiểm tra 0.28 là rất thấp, cho thấy đề thi thiếu sự nhất quán giữa các câu hỏi có thể có câu hỏi không phù hợp, quá nhiều chủ đề hỗn tạp, hoặc độ phân biệt kém.

3.2.1. Tham số của mô hình ứng đáp câu hỏi (IRT)

Hệ số MNSQ (Weighted Fit) trong trường hợp này nằm ở khoảng (0,84-1,16). Dữ liệu kết quả phân tích cho thấy: tất cả 40 câu hỏi đều có hệ số MNSQ nằm trong ngưỡng hợp lý (0.84 – 1.16) chứng tỏ các câu hỏi đều phù hợp với mô hình

Rasch. Các câu hỏi 29 và 34 có giá trị T-statistics < -2 , cần xem lại nội dung, cách diễn đạt, hoặc logic của câu hỏi. Giá trị T-statistics nằm ngoài khoảng khuyến nghị $(-2;2)$ (Paul. NewBold -2013; Moore -2020) cho thấy sự sai lệch của câu hỏi so với mô hình Rasch là rất lớn và có ý nghĩa thống kê, cho thấy câu hỏi đó không phù hợp với mô hình và gây ảnh hưởng tiêu cực đến độ chính xác của kết quả đo lường qua đề kiểm tra thi thử đối với thí sinh.

Căn cứ vào Biểu đồ phân bố tiềm ẩn và ước lượng tham số của mô hình phản hồi (MAP OF LATENT DISTRIBUTIONS AND RESPONSE MODEL PARAMETER ESTIMATES), thu thập được dữ liệu thông tin về đề thi như sau:

Phân bố năng lực của thí sinh (mỗi ký tự “X” đại diện cho 1 thí sinh). Thang đo logit không hiển thị trong hình ảnh, chứng tỏ mức năng lực của thí sinh khá hẹp, chỉ dao động quanh giá trị trên dưới 0 (mức năng lực trên và dưới trung bình).

Phân tích phân bố các câu hỏi trong đề thi môn Pháp luật đại cương theo mức năng lực của thí sinh cho thấy chất lượng đề thi còn nhiều bất cập, đặc biệt trong việc đo lường và phân loại chính xác năng lực người học. Với tổng số 40 câu hỏi, phổ năng lực thí sinh tập trung chủ yếu xung quanh giá trị 0 – tương ứng với mức năng lực trung bình. Tuy nhiên, cấu trúc phân bố độ khó của các câu hỏi lại không tương thích với đặc

điểm năng lực này. Cụ thể, **nhóm 1** gồm 13 câu hỏi (chiếm 32,5%) như các câu 3, 5, 11, 14, 17, 18, 20, 24, 29, 32, 34, 38, 40 được xác định là quá khó, nằm ở vùng đỉnh biểu đồ, vượt xa mức năng lực của tất cả thí sinh, dẫn đến xác suất trả lời đúng gần như bằng không. Ngược lại, **nhóm 2** gồm 7 câu hỏi (chiếm 17,5%) như các câu 4, 15, 16, 19, 25, 35, 39 được đánh giá là quá dễ, nằm dưới cả mức năng lực thấp nhất, khiến xác suất trả lời đúng gần như tuyệt đối với mọi thí sinh. **Nhóm 3** chiếm 40% tổng số câu hỏi, bao gồm các câu 1, 2, 7, 9, 10, 13, 22, 27, 28, 30, 33, 37 là những câu có độ khó thấp hơn mức trung bình, trong khi các câu 23, 26, 31, 36 có độ khó cao hơn mức trung bình. Đáng lưu ý, câu 30 mặc dù rất dễ nhưng lại không có nhiều thí sinh đạt mức năng lực tương ứng, cho thấy sự không tương thích giữa đặc điểm câu hỏi và phân bố năng lực. **Nhóm 4** gồm các câu hỏi 6, 8, 12, 21 (chiếm 10%) có độ khó cao, nằm gần vùng đỉnh biểu đồ, với xác suất trả lời đúng rất thấp, chỉ phù hợp với một tỷ lệ rất nhỏ thí sinh. Như vậy, có thể nhận định rằng đề thi này có chất lượng thấp do tỷ lệ câu hỏi quá khó hoặc quá dễ chiếm tới 60% (24/40 câu), tập trung ở hai cực của biểu đồ năng lực, trong khi phổ năng lực của thí sinh lại chủ yếu phân bố quanh mức trung bình.

3.2.2. Tham số một số câu hỏi có vấn đề về phương án nhiễu và đương công đặc trưng câu hỏi

Giá trị độ khó câu hỏi	Mức ý nghĩa
$b < -2$	Câu hỏi rất dễ
$-2 \leq b < -0,5$	Câu hỏi dễ
$-0,5 \leq b < 0,5$	Câu hỏi trung bình
$0,5 \leq b < 2$	Câu hỏi khó
$b \geq 2$	Câu hỏi rất khó

Mức ý nghĩa độ khó của câu hỏi
(Baker, 2001; Lâm Quang Thiệp, 2011)

Câu hỏi 7: Baker (2001) (The Basics of Item Response Theory) cho rằng ngưỡng độ khó (Item Delta(s)) nên nằm trong khoảng từ -2 đến +2 và được chia 5 bậc: rất dễ, dễ, trung bình, khó, rất khó. Đây là quy ước phổ biến trong nhiều nghiên cứu IRT. Ebel (1965) cho rằng các câu hỏi trong bài thi nên có độ phân biệt (Discrimination) đạt từ 0,3 trở lên. Khi giá trị Pt Bis > 0 : Số lượng TS có năng lực cao lựa chọn nhiều hơn số lượng TS có năng lực thấp; khi giá trị Pt Bis < 0 : Số lượng TS có năng lực thấp lựa chọn nhiều hơn số lượng TS

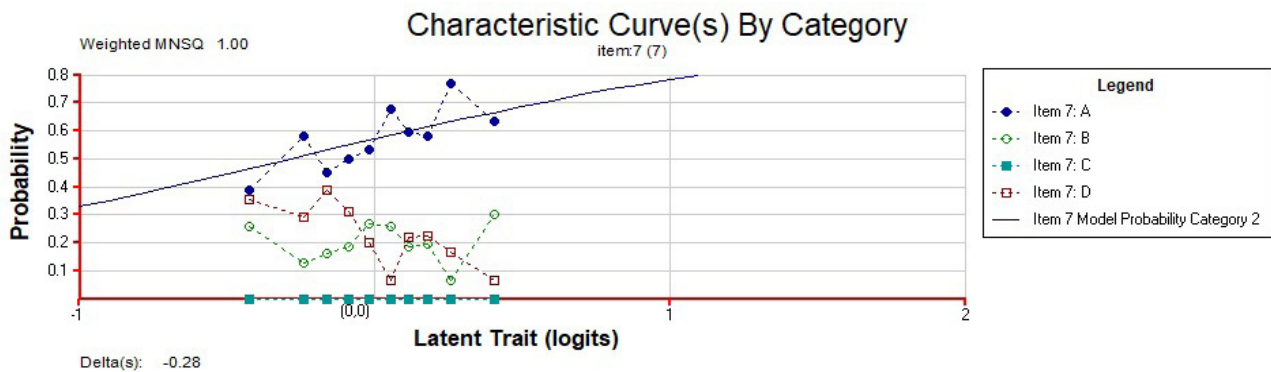
có năng lực cao; khi giá trị $p < 0.05$: chênh lệch giữa số lượng TS lựa chọn các phương án có ý nghĩa thống kê nếu độ tin cậy là 95%. Như vậy, ta kết luận một phương án nhiễu tốt khi giá trị Pt Bis < 0 và giá trị $p < 0.05$. Câu hỏi 7 là câu hỏi dễ (Item Delta(s) = -0,28), đáp án đúng là phương án A (Score=1), phương án nhiễu C có vấn đề vì không có bất cứ thí sinh nào lựa chọn phương án này, không xuất hiện trong phần dữ liệu kết quả của câu hỏi. Độ phân biệt của câu hỏi này rất kém ($a=0,15$).

Giá trị độ phân biệt	Mức ý nghĩa
$a < 0,35$	Độ phân biệt rất kém
$0,35 \leq a < 0,65$	Độ phân biệt kém
$0,65 \leq a < 1,35$	Độ phân biệt trung bình
$1,35 \leq a < 1,7$	Độ phân biệt tốt
$a \geq 1,7$	Độ phân biệt rất tốt

Mức ý nghĩa độ phân biệt của câu hỏi
(Baker, 2001)

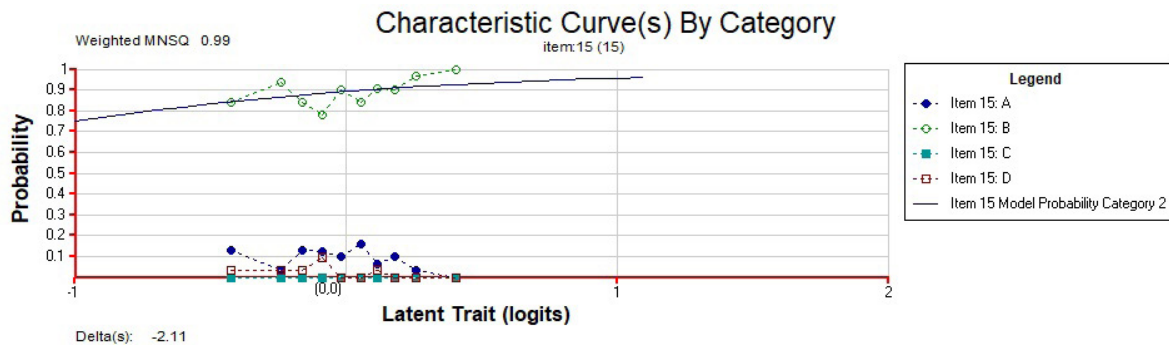
Câu hỏi 15: Độ khó: Câu hỏi quá dễ (độ khó = -2.11), với 89% thí sinh trả lời đúng. Điều này làm giảm khả năng phân biệt giữa các thí sinh, đặc biệt trong một bài thi cần đánh giá toàn diện năng lực. Độ phân biệt: Giá trị 0.25 là quá thấp, cho thấy câu hỏi không phân biệt hiệu quả giữa thí sinh giỏi và kém. Điều này được xác nhận qua Pt Bis (0.25) và sự khác biệt nhỏ về năng lực giữa các nhóm chọn đáp án. Độ phù hợp

mô hình: Weighted MNSQ = 0.99 cho thấy câu hỏi phù hợp với mô hình Rasch, không có dấu hiệu bất thường về mặt kỹ thuật. Các phương án nhiễu A và D không đủ hấp dẫn (8.74% và 2.27%), đặc biệt là đáp án D, cho thấy cần cải thiện các đáp án sai để thu hút nhiều thí sinh hơn và tăng tính thử thách. Điểm bất thường ở câu hỏi này là không có bất cứ thí sinh nào chọn phương án nhiễu C.



Đường cong đặc trưng của câu hỏi số 7 (đáp án đúng: A) Độ khó (delta) là -0.28, cho thấy câu hỏi ở mức dễ. Đáp án đúng A có xác suất chọn cao nhất (đường nét đứt màu xanh tím), tăng dần theo năng lực (Latent Trait). Các phương án nhiễu (B, C, D) có xu hướng đan xen lẫn nhau, thu hút

thí sinh mức năng lực trung bình (xung quanh giá trị 0 logit), xác suất lựa chọn phương án nhiễu C là 0% với mọi thí sinh ở câu hỏi này. Độ phù hợp mô hình (Weighted MNSQ) là 1.00, nằm trong ngưỡng chấp nhận được, cho thấy câu hỏi phù hợp với mô hình Rasch.



Đường cong đặc trưng của câu hỏi số 15 (đáp án đúng B): Độ khó (delta) là -2.11, cho thấy câu hỏi quá dễ, với xác suất chọn đáp án đúng B (đường nét đứt cong xanh lá cây) vượt trội (>80%) ngay ở mức năng lực thấp (dưới giá trị -1 logit). Các đáp

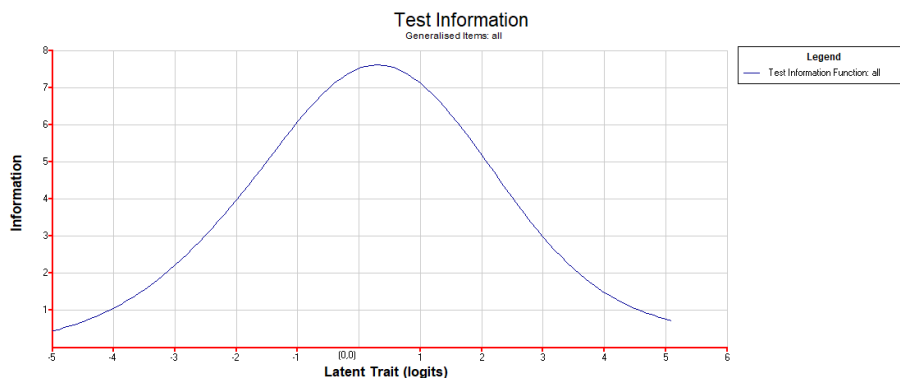
án sai A và D có xác suất chọn rất thấp, chỉ thu hút một số ít thí sinh, làm giảm tính phân biệt của câu hỏi. Độ phù hợp mô hình (Weighted MNSQ) là 0.99, nằm trong ngưỡng chấp nhận được, nhưng độ khó quá thấp và sự áp đảo của đáp án đúng

cho thấy câu hỏi cần cải thiện. Phương án nhiều C cũng không có bất cứ thí sinh nào lựa chọn trong trường hợp này.

3.2.3. Phân tích hàm thông tin đề thi

Trục ngang (Latent Trait - θ , logits): Biểu thị mức độ năng lực của người thi (θ), thường được đo trên thang logits, từ -5 đến +5. Giá trị θ thấp (âm) thể hiện năng lực thấp, giá trị θ cao (dương) thể hiện năng lực cao. Trục dọc (Information): Biểu thị lượng thông tin mà đề thi cung cấp tại một mức năng lực θ . Lượng thông tin càng cao, đề thi càng chính xác trong việc đo lường năng lực tại mức đó. Đường cong (Information Function - all): Là hàm thông tin tổng của đề thi, cho thấy mức độ chính xác của đề thi trong việc đo lường năng lực ở các mức θ khác nhau. Đỉnh của đường cong:

Đỉnh nằm gần $\theta = 0$, với giá trị thông tin khoảng 7.5. Điều này có nghĩa là đề thi cung cấp thông tin tối đa (đo lường chính xác nhất) cho những người có năng lực trung bình ($\theta \approx 0$). Đề thi cung cấp thông tin tốt nhất trong khoảng năng lực (θ) từ -2 đến +2. Ở ngoài khoảng năng lực trên, lượng thông tin giảm mạnh, đề thi không đo lường chính xác đối với thí sinh có năng lực rất thấp ($\theta < -2$) hoặc rất cao ($\theta > 2$). Tại $\theta = -4$ và $\theta = 4$, lượng thông tin gần như bằng 0, cho thấy đề thi gần như không cung cấp thông tin hữu ích đối với thí sinh có năng lực cực thấp hoặc cực cao. Tại $\theta = -1$ và $\theta = 1$, lượng thông tin vẫn khá cao (khoảng 4-5), nghĩa là đề thi vẫn đo lường tốt đối với nhóm thí sinh có năng lực tiệm cận dưới hoặc tiệm cận trên trung bình.



III. KẾT LUẬN

Kết quả phân tích dữ liệu đề thi môn Pháp luật đại cương cho thấy điểm trung bình của thí sinh rất thấp (4,4/10), phản ánh nhiều vấn đề: sự chuẩn bị kiến thức chưa tốt của thí sinh, thái độ xem nhẹ môn học, giảng dạy chưa hiệu quả, đề thi chưa đạt chuẩn. Đề thi có đến 60% câu hỏi quá dễ hoặc quá khó, độ tin cậy thấp (Cronbach's Alpha = 0,28), cho thấy cần cải thiện toàn diện về cấu trúc và

nội dung câu hỏi nhằm đảm bảo đo lường chính xác năng lực người học. Trường cần áp dụng các nguyên tắc thiết kế đề thi đảm bảo tính khách quan, độ giá trị, độ tin cậy và nên thử nghiệm trước khi thi chính thức. Việc ứng dụng hệ thống trắc nghiệm thích ứng như UEd-CAT 1.0 có thể nâng cao chất lượng đánh giá, tiết kiệm thời gian, tăng tính bảo mật và đánh giá chính xác năng lực thí sinh.

TÀI LIỆU THAM KHẢO

- Lê, T. H., Nguyễn, Q. T., Nguyễn, T. Q. G., & Nguyễn, T. H. (2024). Phương pháp kiểm tra đánh giá năng lực người học bằng bài thi trắc nghiệm thích ứng trên máy tính (CAT) (Bằng độc quyền sáng chế số 40296). Cục Sở hữu trí tuệ, Bộ Khoa học và Công nghệ.
- Lê Thái Hưng, Nguyễn Văn Tuấn, Dương Thị Anh (2015), Xây dựng ngân hàng câu hỏi đánh giá kết quả học tập môn vật lý học kì 1 lớp 12: vận dụng IRT và phần mềm conquest, tạp chí Quản lý giáo dục số 79.12/2015, 40 – 46
- Sái Công Hồng - Lê Thái Hưng - Lê Thị Hoàng Hà - Lê Đức Ngọc, (2017), Giáo trình Kiểm tra Đánh giá trong dạy học, NXB Đại học Quốc gia Hà Nội.
- Nguyễn, T. H., Bùi, T. K. P., & Lê, T. H. (2023). Phát triển ngân hàng câu hỏi trắc nghiệm thích ứng đánh giá từ vựng tiếng Anh thông dụng: áp dụng IRT và phương pháp cân bằng đề. Tạp Chí Giáo dục, 23(19), 8-14. Truy vấn từ <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/1066>
- Nguyễn Quỳnh Giang, Lê Thái Hưng (2018). Mô phỏng một bài kiểm tra thích nghi trên máy tính thông qua phần mềm R. Tạp chí Khoa học Giáo dục Việt Nam, 11, 6-11.
- Bui, T. K. P., Nguyen, Q. T., & Le, T. H. (2022). The development of A Vietnamese-English Bilingual Version of the New General Service List Test. 2nd Hanoi Forum on Pedagogical and Educational Sciences, Hanoi, Vietnam
- Le, T. H., & Nguyen, T. H. (2021). Experimental Research and Application of Computerized Adaptive Tests to assess

Learners' Competencies. 2021 3rd International Conference on Computer Science and Technologies in Education (CSTE), Beijing, China. <https://doi.org/10.1109/CSTE53634.2021.00021>

Adams, R. J., & Wu, M. L. (2007). The mixed-coefficients multinomial logit model: A generalized form of the Rasch model. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 57–75). Springer. https://doi.org/10.1007/978-0-387-49839-3_5

Bộ Giáo dục và Đào tạo. (1995). Quyết định số 3244/GD-ĐT ngày 12 tháng 9 năm 1995 về việc đưa môn Pháp luật đại cương vào giảng dạy chính thức trong các trường đại học, cao đẳng và trung cấp chuyên nghiệp.

Quang, B. N. (2017). Đánh giá chất lượng ngân hàng đề thi trắc nghiệm khách quan môn Nhân học đại cương bằng mô hình Rasch và phần mềm Quest. *SCIENCE & TECHNOLOGY*, 20(X3-2017).

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>

Ebel, R. L. (1965). *Measuring educational achievement*. Prentice-Hall.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer Academic Publishers.

Khuyến, N. T. (2017). Đổi mới phương pháp giảng dạy học phần pháp luật đại cương tại Trường Đại học Tiền Giang. *Tạp chí Khoa học Đại học Đồng Tháp*, (25), 42–47.

Lâm Quang Thiệp. (2011). *Đo lường trong giáo dục: Lý thuyết và Ứng dụng*. NXB Đại học Quốc gia Hà Nội.

Moore, D. S., Notz, W. I., & Fligner, M. A. (2020). *The basic practice of statistics* (9th ed.). W. H. Freeman.

Newbold, P., Carlson, W. L., & Thorne, B. (2013). *Statistics for business and economics* (8th ed.). Pearson.

Nguyễn, V. C., & Nguyễn, Q. T. (2021). Sử dụng phần mềm Quest, Conquest để phân tích đề thi tự luận. *Tạp chí Khoa học Đại học Đồng Tháp*, (18), 52–64.

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.

Phan, T. D. T. N. T., & Hiền, T. (2019). *Pháp luật đại cương*.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. MESA Press.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. ACER Press.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.