

INITIAL EVALUATION OF THE MCQ QUESTION BANK ON THE CELL CYCLE AND CELL DIVISION USING THE TWO-PARAMETER IRT MODEL

Cao Thi Tai Nguyen

Can Tho University of Medicine and Pharmacy

| ARTICLE INFO | ABSTRACT |
|--|---|
| <p>Received: 13/01/2025</p> <p>Revised: 17/02/2025</p> <p>Published: 19/02/2025</p> | <p>The two-parameter Item Response Theory model has been widely used to assess the quality of test questions across various subjects. However, its application in evaluating regular tests in Genetics-Biology in Vietnam is still limited. The aim of this study is to evaluate the quality of multiple-choice questions in regular tests by determining the difficulty and discrimination indices of the questions and identifying the proportion of questions that need to be revised or removed. The study used the two-parameter IRT model, with data collected from 480 regular tests containing 119 questions from students of Cohorts 47 and 35, enrolled in the lesson on Cell Cycle and Cell Division of Biology - Genetics course at Can Tho University of Medicine and Pharmacy during the third semester of the 2021-2022 academic year. The analysis was conducted using the "lrm" package in R. The results showed that 82% of the questions were easy, 13% were of average difficulty, and 4.2% were very easy. Regarding discrimination, 42% of the questions had very good discrimination, 33% were good, and 20% were average. However, 5% of the questions showed poor discrimination and need to be revised. The conclusion highlights that the two-parameter IRT model is an effective tool for assessing the quality of questions in a question bank. Most of the questions (113/119) meet the requirements for difficulty and discrimination, making them suitable for inclusion in the question bank. It is also necessary to add questions of medium, difficult, and very difficult levels to ensure the test matrix has a sufficient range of difficulty levels.</p> |
| <p>KEYWORDS</p> <p>Item Response Theory</p> <p>Two-parameter IRT model</p> <p>Regular test</p> <p>MCQ</p> <p>R software</p> | |

BƯỚC ĐẦU ĐÁNH GIÁ BỘ CÂU HỎI MCQ BÀI CHU KỲ TẾ BÀO VÀ SỰ PHÂN CHIA TẾ BÀO BẰNG MÔ HÌNH IRT 2 THAM SỐ

Cao Thị Tài Nguyên

Trường Đại học Y Dược Cần Thơ

| THÔNG TIN BÀI BÁO | TÓM TẮT |
|---|---|
| <p>Ngày nhận bài: 13/01/2025</p> <p>Ngày hoàn thiện: 17/02/2025</p> <p>Ngày đăng: 19/02/2025</p> | <p>Mô hình lý thuyết ứng đáp câu hỏi 2 tham số đã được sử dụng rộng rãi để đánh giá chất lượng câu hỏi kiểm tra, nhưng ứng dụng trong đánh giá bài kiểm tra Sinh học - Di truyền ở Việt Nam còn hạn chế. Mục tiêu nghiên cứu này là đánh giá chất lượng câu hỏi trắc nghiệm qua tỷ lệ độ khó, độ phân biệt và tỷ lệ câu hỏi cần chỉnh sửa hoặc loại bỏ. Dữ liệu nghiên cứu gồm 480 bài kiểm tra với 119 câu hỏi của sinh viên khóa 47 và 35, tham gia bài Chu kỳ tế bào và sự phân chia tế bào trong khóa học môn Sinh học - Di truyền tại Trường Đại học Y Dược Cần Thơ ở học kỳ 3 năm học 2021-2022. Nghiên cứu sử dụng gói "lrm" trong R để phân tích dữ liệu. Kết quả cho thấy, 82% câu hỏi dễ, 13% ở mức trung bình, 4,2% rất dễ. Về độ phân biệt, 42% câu hỏi phân biệt rất tốt, 33% tốt, 20% bình thường, 5% câu hỏi có độ phân biệt kém và cần chỉnh sửa. Kết luận cho thấy mô hình lý thuyết ứng đáp câu hỏi 2 tham số là công cụ hiệu quả trong đánh giá chất lượng câu hỏi trong ngân hàng câu hỏi. Đa số câu hỏi (113/119) đáp ứng yêu cầu về độ khó và độ phân biệt, có thể sử dụng trong ngân hàng câu hỏi. Tuy nhiên, cần bổ sung câu hỏi có độ khó trung bình, khó và rất khó để đảm bảo độ khó và phân biệt cho bài thi, tạo ma trận đề thi đầy đủ các mức độ yêu cầu.</p> |
| <p>TỪ KHÓA</p> <p>Lý thuyết ứng đáp câu hỏi</p> <p>Mô hình IRT 2 tham số</p> <p>Kiểm tra thường xuyên</p> <p>MCQ</p> <p>Phần mềm R</p> | |

DOI: <https://doi.org/10.34238/tnu-jst.11873>

Email: ctnguyen@ctump.edu.vn - tainguyen77ct@gmail.com

<http://jst.tnu.edu.vn>

160

Email: jst@tnu.edu.vn

1. Mở đầu

Trong lĩnh vực đo lường giáo dục, việc áp dụng lý thuyết ứng đáp câu hỏi (Item Response Theory - IRT) ngày càng phổ biến nhờ khả năng phân tích đặc tính câu hỏi và đánh giá năng lực người học chính xác hơn so với các phương pháp truyền thống. Đặc biệt, mô hình logistic hai tham số (2PL - Two-Parameter Logistic Model) của IRT được sử dụng rộng rãi để đo lường độ khó và khả năng phân biệt của các câu hỏi trắc nghiệm khách quan (MCQ) [1] - [3]. Mô hình này đóng vai trò quan trọng trong việc phát triển các bài kiểm tra chuẩn hóa, nâng cao tính công bằng và đảm bảo độ tin cậy của các đánh giá [4] - [7]. Độ khó của một câu hỏi thể hiện mức độ mà người trả lời cần phải có khả năng cao để trả lời đúng. Trong mô hình IRT, nó được ký hiệu là b_j , là điểm trên thang năng lực nơi xác suất trả lời đúng là 50%. Ở đây, b_j là giá trị năng lực (θ) mà người trả lời cần đạt được để có xác suất trả lời đúng là 50%. Năng lực của người trả lời (θ) được đo trên một thang chuẩn hóa, với giá trị trung bình thường là 0 và độ lệch chuẩn là 1. Chẳng hạn, câu hỏi với $b_j = 1,5$ yêu cầu người trả lời có năng lực 1,5 (trên thang đo chuẩn hóa) để có xác suất 50% trả lời đúng. Độ phân biệt của một câu hỏi thể hiện khả năng phân biệt giữa những người có năng lực khác nhau. Trong mô hình IRT, nó thường được ký hiệu là a_j . Câu hỏi có độ phân biệt cao (giá trị a_j lớn) sẽ rõ ràng hơn trong việc phân biệt người có năng lực cao và năng lực thấp, thường có đường cong đặc trưng mục kiểm tra (Item Characteristic Curve - ICC) dốc hơn.

Mô hình IRT 2 tham số được mô tả bởi hai tham số chính: a_j (độ phân biệt) và b_j (độ khó). Tham số a_j thể hiện khả năng của câu hỏi trong việc phân biệt giữa người có năng lực cao và thấp, trong khi tham số b_j biểu thị mức độ khó của câu hỏi. Xác suất trả lời đúng một câu hỏi trong mô hình này được xác định bởi hàm:

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

Trong đó: θ là năng lực của người trả lời và a, b là tham số đặc trưng của câu hỏi.

So với mô hình 1 tham số (1PL), mô hình 2 tham số bổ sung tham số a_j , giúp phản ánh độ phân biệt của câu hỏi, từ đó cung cấp đánh giá chi tiết và chính xác hơn về chất lượng câu hỏi.

Mô hình 2PL cung cấp hiểu biết sâu sắc về cách từng câu hỏi trong bài kiểm tra hoạt động đối với các nhóm năng lực khác nhau của người học [8], [9]. Nhiều nghiên cứu đã chứng minh hiệu quả của mô hình này trong các ứng dụng như kiểm tra y khoa [2], hệ thống kiểm tra thích nghi [7] và tâm lý học giáo dục [5], [10]. Ngoài ra, mô hình 2PL cũng hỗ trợ tối ưu hóa cấu trúc bài kiểm tra và lựa chọn câu hỏi phù hợp trong môi trường kiểm tra dựa trên máy tính [7], [8].

Tuy nhiên, việc triển khai các mô hình IRT, bao gồm mô hình 2PL, đòi hỏi công cụ tính toán tiên tiến và sự am hiểu chuyên sâu về phân tích đo lường. Các công cụ như phân mềm R và Quest/ConQuest đã được chứng minh là hỗ trợ hiệu quả cho việc phân tích dựa trên IRT [5], [8], [11]. Cùng với đó, những tiến bộ gần đây đã phát triển các thuật toán kiểm tra thích nghi, cho phép bài kiểm tra tự động điều chỉnh theo năng lực của người học, chứng minh tính linh hoạt của IRT trong cả định dạng đánh giá nhị phân và đa cấp [8], [9], [12].

Ở Việt Nam, IRT không chỉ được áp dụng trong giáo dục mà còn trong các lĩnh vực y tế, tâm lý học và nghiên cứu xã hội. Trong giáo dục, IRT đóng vai trò quan trọng trong việc phát triển và đánh giá các bài kiểm tra trắc nghiệm. Các nhà nghiên cứu như Nguyễn Văn Cảnh [3] và Phạm Văn Tác [4] đã áp dụng các mô hình IRT từ 1 đến 3 tham số nhằm cải thiện chất lượng đo lường và đánh giá giáo dục. Những nỗ lực này đã góp phần xây dựng các bài kiểm tra đạt tiêu chuẩn, bảo đảm tính khách quan và độ tin cậy cao. Trong lĩnh vực tâm lý học, IRT được sử dụng để thiết kế và chuẩn hóa các bài kiểm tra đo lường các đặc điểm tâm lý như mức độ hài lòng, lo âu và trầm cảm. Những ứng dụng này đã khẳng định vai trò quan trọng của IRT trong việc nâng cao chất lượng các công cụ đo lường ở nhiều lĩnh vực tại Việt Nam. Tuy nhiên mô hình này hiện vẫn chưa được sử dụng trong đánh giá câu hỏi MCQ tại trường Đại học Y Dược Cần Thơ.

Mô hình IRT giúp xác định độ khó và độ phân biệt của từng câu hỏi, cung cấp thông tin quan trọng cho việc điều chỉnh chất lượng ngân hàng câu hỏi để thiết kế đề thi đảm bảo tính công bằng

và đồng nhất. Ngân hàng câu hỏi theo quy định của Trường Đại học Y Dược Cần Thơ là tối thiểu 100 câu cho 1 tiết học. Bộ môn đã xây dựng được 3000 câu cho 2 tín chỉ với 30 tiết của môn Sinh học Di truyền. Tuy nhiên, hiện tại, sự phân loại độ khó của câu hỏi thường chỉ được thực hiện một cách chủ quan từ giảng viên của bộ môn. Để nâng cao chất lượng ngân hàng câu hỏi, việc đánh giá dựa trên kết quả của người học trở nên cần thiết hơn bao giờ hết. Bên cạnh đó, việc đánh giá chất lượng câu hỏi trong ngân hàng câu hỏi, đặc biệt là ngân hàng câu hỏi môn Sinh học Di truyền cho nội dung về Chu kỳ tế bào và sự phân chia tế bào ở các trường đại học, cũng cần được thực hiện một cách khoa học và hệ thống để đảm bảo tính khách quan và đồng nhất.

Mô hình IRT là một công cụ mạnh mẽ trong việc đánh giá độ khó và khả năng phân biệt của các câu hỏi, cũng như đo lường năng lực của sinh viên. Trong phạm vi của việc đánh giá chất lượng câu hỏi về Chu kỳ tế bào và sự phân chia tế bào trong ngân hàng câu hỏi môn Sinh học - Di truyền, nghiên cứu này sử dụng mô hình IRT 2 tham số. Mục tiêu của nghiên cứu này là áp dụng mô hình này để đánh giá chất lượng ngân hàng câu hỏi, bao gồm việc xác định độ khó và độ phân biệt của từng câu hỏi trắc nghiệm thông qua bài kiểm tra thường xuyên. Đồng thời, nghiên cứu cũng nhằm xác định tỷ lệ câu hỏi chưa đạt yêu cầu, từ đó đề xuất các điều chỉnh hoặc loại bỏ những câu hỏi này khỏi ngân hàng câu hỏi của môn học Chu kỳ tế bào và sự phân chia tế bào trong ngân hàng câu hỏi môn Sinh học - Di truyền tại Trường Đại học Y Dược Cần Thơ. Điều này nhằm đảm bảo chất lượng và đồng nhất của đề thi và ngân hàng câu hỏi, từ đó tạo điều kiện thuận lợi cho quá trình học tập và đánh giá của sinh viên.

2. Phương pháp nghiên cứu

Trong quá trình giảng dạy, việc đánh giá kiểm tra thường xuyên là cần thiết trong quá trình đánh giá môn học của mỗi sinh viên. Kiểm tra thường xuyên chiếm 30% trong số tổng điểm thành phần của môn học. Kiểm tra thường xuyên là một hoạt động thường quy được đánh giá qua câu hỏi MCQ tại bộ môn Sinh học - Di truyền. Ở học kỳ 3 năm học 2021-2022, 480 sinh viên khoá 47 và khoá 35 đã hoàn thành bài kiểm tra thường xuyên online bài Chu kỳ tế bào và sự phân chia tế bào môn Sinh học - Di truyền. Hình thức kiểm tra thường xuyên là làm trên link online của Microsoft form được gửi cho sinh viên với tài khoản chuyên biệt cho từng sinh viên và mỗi tài khoản chỉ có 1 lần làm bài. Link được gửi vào 20h00 - 21h30 ngày 20 tháng 6 năm 2022. Thời gian thu bài được cài là 21h30 theo giờ GMT của thế giới. Sinh viên tự thực hiện làm bài kiểm tra thường xuyên trên thiết bị di động hoặc laptop hoặc máy tính cá nhân có kết nối wifi. Trong 119 câu MCQ trên link được gửi cho sinh viên làm bài thì mỗi tài khoản sinh viên thực hiện làm bài đều bị xáo trộn câu hỏi và mỗi nhữ nhằm hạn chế việc trao đổi bài và quay cóp vì hết thời gian quy định sinh viên không thể nộp được bài nếu chưa hoàn thành việc trả lời 119 câu. Hệ thống không tự thu bài và nếu có bất kỳ câu nào chưa trả lời đều không nộp được bài. Trong quá trình nộp bài nếu sinh viên nào có sự cố về đường truyền thì được mở thêm thời gian 5 phút để các bạn có thể nộp bài. Trong thời gian này nếu các bạn không nộp được thì xem như nộp bài không thành công. Kết quả của toàn bộ sinh viên Khoá 47 và khoá 35 tham gia học ở học kỳ 3 nộp bài thành công được lấy để phân tích từng câu hỏi trong đề kiểm tra thường xuyên. Đề kiểm tra gồm có 119 câu hỏi trắc nghiệm khách quan 4 lựa chọn với 1 đáp án đúng. Dữ liệu câu trả lời trên file Excel với 480 sinh viên nộp bài thành công và lấy cột kết quả của mỗi câu đã mã hóa dữ liệu dạng nhị phân 1 và 0. Sinh viên trả lời đúng thì đã được hệ thống gán giá trị 1, ngược lại được gán giá trị 0.

File Excel được đưa vào R và dùng gói lệnh “lrm” để phân tích. Gói lệnh này chuyên được dùng để đo lường độ khó, độ phân biệt của các câu hỏi trong bài kiểm tra thường xuyên. Nhập dữ liệu vào R, và sử dụng mô hình IRT 2 tham số để ước lượng độ khó, độ phân biệt của mỗi câu hỏi trong bài kiểm tra thường xuyên. Trong gói “lrm” của R, các thông số b_j (độ khó) và a_j (độ phân biệt) của các câu hỏi được ước tính thông qua phương pháp cực đại hóa hợp lý (Maximum Likelihood Estimation - MLE) trong khuôn khổ của mô hình IRT. Cụ thể, gói lrm sử dụng mô hình logistic 2 tham số (2PL) với yêu cầu một ma trận dữ liệu nhị phân (0/1), trong đó dữ liệu

chứa giá trị mỗi hàng là câu trả lời của một người trả lời và mỗi cột là một câu hỏi/mục kiểm tra. Các thông số a_j và b_j được tính toán bằng cách tối đa hóa hàm hợp lý của dữ liệu.

Căn cứ vào các giá trị về độ khó (b_j) và độ phân biệt (a_j), chất lượng các câu hỏi được phân loại và đánh giá dựa theo các thang đo của Baker. Độ khó của các câu hỏi theo 5 mức sau: rất khó, khó, trung bình, dễ, rất dễ. Một câu hỏi thuộc loại rất khó nếu tham số $b_j \geq 2$, thuộc loại khó nếu $1 \leq b_j < 2$, thuộc loại trung bình nếu $-1 \leq b_j < 1$, thuộc loại dễ nếu $-2 < b_j < -1$ và thuộc loại rất dễ nếu $b_j \leq -2$. Độ phân biệt của các câu hỏi gồm 5 mức: rất tốt, tốt, bình thường, kém và rất kém. Cụ thể, một câu hỏi được gọi là có độ phân biệt rất tốt nếu tham số $a_j \geq 1,7$, loại tốt nếu $1,35 \leq a_j \leq 1,69$, loại bình thường nếu $0,65 \leq a_j \leq 1,34$, loại kém nếu $0,35 \leq a_j \leq 0,64$ và loại rất kém nếu $a_j \leq 0,34$ [13].

3. Kết quả và bàn luận

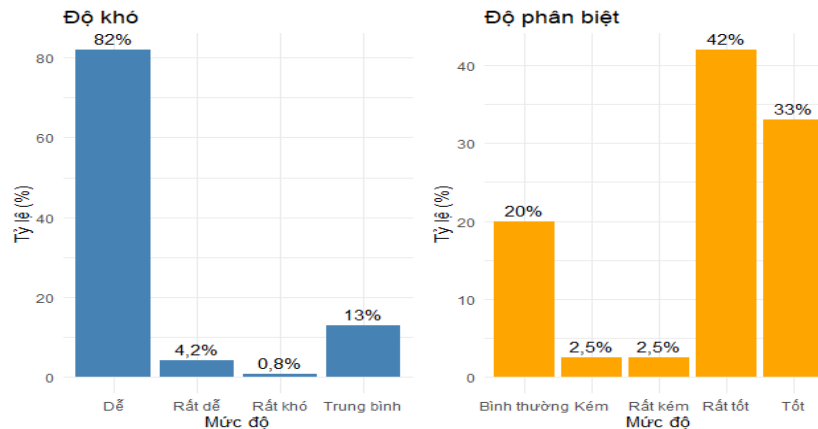
Bảng gói lệnh "lrm" trong R với mô hình 2 tham số của IRT, kết quả ghi nhận các giá trị của cột Dffclt chỉ độ khó của các câu hỏi và cột cuối Dscrmm chỉ độ phân biệt của các câu hỏi. Hình 1 thể hiện kết quả phân tích của các câu hỏi trong bài kiểm tra thường xuyên MCQ với giá trị độ khó và độ phân biệt.

| | Dffclt | Dscrmm |
|--------|------------|-------------|
| Item1 | -1.6599800 | 1.87330045 |
| Item2 | -1.4742374 | 1.13291459 |
| Item3 | -1.9294284 | 1.93532991 |
| Item4 | -1.8659810 | 2.13680901 |
| Item5 | -1.5351544 | 1.24688136 |
| Item6 | -1.7171071 | 1.80001726 |
| Item7 | -1.1403343 | 1.08399419 |
| Item8 | -1.0227357 | 1.48670335 |
| Item9 | -0.8058780 | 1.24029675 |
| Item10 | -1.1160150 | 1.73087911 |
| Item11 | -1.5402326 | 0.78360967 |
| Item12 | -0.4790663 | 1.35453563 |
| Item13 | -1.8949828 | 0.82309462 |
| Item14 | -1.4196707 | 1.93792213 |
| Item15 | 0.3346486 | -0.45478779 |

Hình 1. Kết quả phân tích với độ khó và độ phân biệt của bài kiểm tra thường xuyên

Từ dữ liệu phân tích về độ khó độ phân biệt, nghiên cứu ghi nhận về tỷ lệ độ khó câu hỏi trong bài kiểm tra thường xuyên như sau: có 1 (0,8%) câu hỏi rất khó, 0 (0%) câu hỏi khó, 15 (13%) câu hỏi ở mức trung bình, 98 (82%) câu hỏi dễ và 5 (4,2%) câu hỏi rất dễ (Hình 2). Kết quả này có sự khác biệt so với nghiên cứu của Adetutu và Lawal [13] cũng như Đoàn Hồng Chương và cộng sự [14]. Trong nghiên cứu [13], 49% câu hỏi trong bài kiểm tra được đánh giá là rất dễ, trong khi chỉ có một tỷ lệ nhỏ các câu hỏi có độ khó cao (2,9%). Nghiên cứu đánh giá trên 35 câu hỏi trắc nghiệm 4 lựa chọn với 403 sinh viên tham gia kỳ thi môn thống kê năm 2017-2018, cho thấy phần lớn câu hỏi tập trung vào độ khó thấp và trung bình. Tương tự, nghiên cứu của Đoàn Hồng Chương [14] khảo sát trên 388 sinh viên thi môn Toán cao cấp với 20 câu hỏi MCQ 4 lựa chọn chỉ ghi nhận có 5/20 (25%) câu hỏi rất dễ, 8/20 (40%) câu hỏi dễ và 7/20 (35%) câu hỏi ở mức độ trung bình. Mặc dù tỷ lệ câu hỏi có độ khó trong nghiên cứu của chúng tôi có sự khác biệt, nhưng cũng có thể lý giải rằng mục đích đánh giá của chúng tôi khác với các nghiên cứu trước đó. Cụ thể, trong nghiên cứu của chúng tôi, mục tiêu là đánh giá khả năng cơ bản của sinh viên trong một bài kiểm tra thường xuyên, thay vì đánh giá khả năng áp dụng kiến thức ở mức độ khó hơn.

Độ phân biệt của các câu hỏi, bài kiểm tra thường xuyên có 50/119 (42%) câu hỏi ở mức phân biệt rất tốt, 39/119 (33%) câu hỏi ở mức tốt, 24/119 (20%) câu hỏi ở mức bình thường, 03/119 (2,5%) câu hỏi ở mức kém và 3/119 (2,5%) câu hỏi ở mức rất kém, trong đó có 1 câu có độ phân biệt âm ở nhóm này (câu 15) (Hình 2). Kết quả này có sự khác biệt rõ rệt so với nghiên cứu [13] và [14], đặc biệt là tỷ lệ câu hỏi có độ phân biệt ở các nhóm kém và rất kém. Trong nghiên cứu [13], nhóm câu hỏi có độ phân biệt từ bình thường đến tốt và rất tốt chỉ chiếm tỷ lệ 51,4% (18/35), trong khi nhóm câu hỏi có độ phân biệt kém và rất kém chiếm tỷ lệ cao (31,4% và 11,4%). Điều này cho thấy, trong nghiên cứu [13], tỷ lệ câu hỏi có độ phân biệt thấp khá cao, điều này có thể ảnh hưởng đến khả năng phân biệt nhóm sinh viên có năng lực khác nhau. So với nghiên cứu [14] trên 20 câu hỏi MCQ, tỷ lệ câu hỏi ở nhóm rất tốt chiếm 30% (6/20), nhóm tốt là 5% (1/20), nhóm bình thường 30% (6/20), nhóm kém là 20% (4/20) và nhóm rất kém chiếm 15% (3/20). Tỷ lệ câu hỏi có độ phân biệt kém và rất kém trong nghiên cứu của chúng tôi là 5%, thấp hơn so với nghiên cứu [14] (35%). Như vậy, tỷ lệ câu hỏi phân biệt tốt và rất tốt trong bài kiểm tra của chúng tôi là khá cao, cho thấy bài kiểm tra có khả năng phân biệt rõ rệt giữa các nhóm sinh viên.



Hình 2. Tỷ lệ phân bố độ khó và độ phân biệt của 119 câu hỏi MCQ

Trong nhóm câu hỏi rất dễ có 3 trên 5 câu có thể đưa vào ngân hàng vì có 2 câu có độ phân biệt bình thường và 1 câu có độ phân biệt rất tốt. Trong nhóm câu hỏi dễ có 97 trên 98 câu có thể đưa vào ngân hàng. Trong nhóm câu hỏi trung bình có 13 trên 15 câu có thể đưa vào ngân hàng. Trong bài kiểm tra thường xuyên không có câu hỏi nào có độ khó ở mức khó. Một câu hỏi ở mức rất khó nhưng không đạt được độ phân biệt, nên cần chỉnh sửa hoặc loại bỏ câu hỏi này ra khỏi ngân hàng (Bảng 1).

Bảng 1. Sự phân bố độ khó và độ phân biệt của 119 câu hỏi trong bài kiểm tra thường xuyên

| Độ phân biệt | Độ khó | | | |
|--------------|---------------|------------|--------------------|----------------|
| | Rất dễ, N = 5 | Dễ, N = 98 | Trung bình, N = 15 | Rất khó, N = 1 |
| Rất kém | 1 (20%) | 0 (0%) | 1 (6,7%) | 1 (100%) |
| Kém | 1 (20%) | 1 (1,0%) | 1 (6,7%) | 0 (0%) |
| Bình thường | 2 (40%) | 15 (15%) | 7 (47%) | 0 (0%) |
| Tốt | 0 (0%) | 34 (35%) | 5 (33%) | 0 (0%) |
| Rất tốt | 1 (20%) | 48 (49%) | 1 (6,7%) | 0 (0%) |

Như vậy, bài kiểm tra thường xuyên có 1 trên 119 (0,8%) câu hỏi rất khó, 0 (0%) câu hỏi khó, 15 trên 119 (13%) câu hỏi ở mức trung bình, 98 trên 119 (82%) câu hỏi dễ và 5 trên 119 (4,2%) câu hỏi rất dễ. Về độ phân biệt của các câu hỏi, bài kiểm tra thường xuyên có 50 trên 119 (42%) câu hỏi ở mức phân biệt rất tốt, 39 trên 119 (33%) câu hỏi ở mức tốt, 24 trên 119 (20%) câu hỏi ở mức bình thường, 3 trên 119 (2,5%) câu hỏi ở mức kém và 3 trên 119 (2,5%) câu hỏi ở mức rất kém. Kết quả cho thấy bài kiểm tra thường xuyên chủ đề này chủ yếu câu hỏi đạt về độ khó và độ phân biệt có thể đưa vào ngân hàng là 113 trên 119 câu (95%). Cụ thể số câu hỏi ở mức độ dễ là 97 trên 119, ở mức trung bình có 13 trên 119 và 3 trên 119 câu ở mức rất dễ. Nghiên cứu ghi nhận có 6 trên 119 câu (5,0%) (câu 15, 27, 34, 68, 95 và 118) trong bài kiểm tra thường xuyên là các câu hỏi có độ phân biệt kém và rất kém. Các câu hỏi khó và rất khó đều không đạt chất lượng và đúng chuẩn của một câu hỏi theo phân loại của Baker về độ phân biệt. Chi tiết độ khó và độ phân biệt của 6 câu được thể hiện ở Bảng 2.

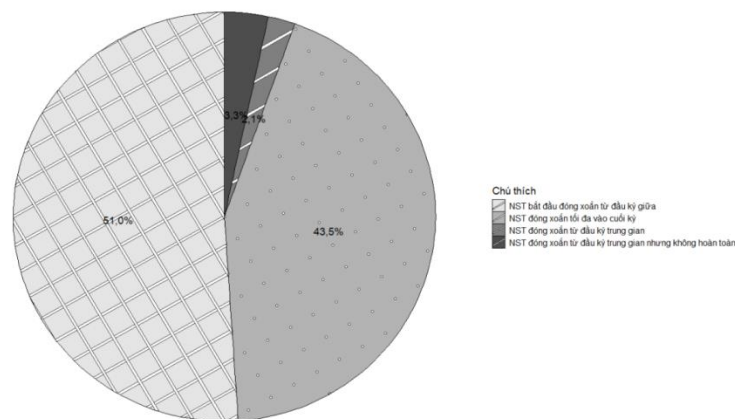
Bảng 2. Các câu hỏi trong bài kiểm tra thường xuyên có độ phân biệt kém và rất kém

| Câu hỏi | Độ khó | Độ phân biệt | Nhóm độ khó | Nhóm độ phân biệt |
|---------|--------------|--------------|-------------|-------------------|
| 15 | 0,334648616 | -0,454787786 | Trung bình | Rất kém |
| 27 | 8,585575151 | 0,041380952 | Rất khó | Rất kém |
| 68 | -2,829651421 | 0,327057014 | Rất dễ | Rất kém |
| 34 | -3,018889087 | 0,540538337 | Rất dễ | Kém |
| 95 | 0,772993336 | 0,388468029 | Trung bình | Kém |
| 118 | -1,342571099 | 0,438649787 | Dễ | Kém |

Chẳng hạn như các câu có độ phân biệt kém ở Hình 3 thì câu 15 (Item15) có độ phân biệt có giá trị âm (độ phân biệt rất kém) và độ khó ở mức trung bình. Nội dung câu 15 từ bài kiểm tra

thường xuyên và kết quả chọn câu đúng của sinh viên được thể hiện ở Hình 3. Chỉ có 51% (245/480) sinh viên chọn đúng. Có 209 trên 480 (43,5%) sinh viên chọn vào mỗi thứ 2 từ trên xuống trong chú thích ở Hình 3.

Phát biểu đúng khi mô tả sự biến đổi hình thái NST qua nguyên phân:



Hình 3. Các câu hỏi có độ phân biệt kém và rất kém trong bài kiểm tra thường xuyên

Nhìn chung, tỷ lệ câu hỏi đạt chuẩn về độ khó và độ phân biệt trong bài kiểm tra thường xuyên của chúng tôi chiếm 95% (113 trên 119) và có thể đưa vào ngân hàng để sử dụng. Tuy nhiên, chủ yếu là câu hỏi có độ khó ở mức dễ (97 trên 113). Trong bài kiểm tra này chưa có câu hỏi nào có độ khó là khó và rất khó mà đạt được độ phân biệt. Như vậy, bài kiểm tra thường xuyên là dễ để sinh viên đạt điểm cao ở nội dung này. Kết quả đánh giá câu hỏi này cần đưa vào dữ liệu ngân hàng câu hỏi của bộ môn để việc đánh giá độ khó câu hỏi không còn là ý kiến chủ quan của giảng viên trong công tác ra đề thi như đã từng được thực hiện. Các câu hỏi cần được xem xét chỉnh sửa hoặc loại bỏ khỏi bài kiểm tra thường xuyên trong ngân hàng do độ phân biệt kém và rất kém gồm câu 15, 27, 34, 68, 95 và 118. Từ kết quả cho thấy 4 trên 6 câu có độ phân biệt kém đều nằm ở nội dung liên quan đến việc dẫn xoắn nhiễm sắc thể (NST) ở kỳ sau hay ở kỳ cuối. Việc sử dụng từ chưa rõ ràng trong đáp án hay trong câu dẫn có thể gây hiểu nhầm cho sinh viên. Cụ thể ở đây là “bắt đầu dẫn xoắn” hay “dẫn xoắn” làm cho sinh viên dễ bị nhầm lẫn và dẫn đến sinh viên có thể hiểu sai ý câu hỏi hoặc không hiểu câu hỏi chính xác. Vì vậy, họ có thể chọn đáp án theo cách ngẫu nhiên hoặc chọn sai đáp án. Với các câu hỏi dễ, sinh viên thường sẽ chọn câu trả lời đúng mà không cần dự đoán, nhưng với câu hỏi có độ phân biệt kém và rất kém, sinh viên dường như gặp khó khăn. Điều này đã khẳng định sự tồn tại của lỗi thiết kế trong câu hỏi này, và việc sửa đổi hoặc loại bỏ câu hỏi này có thể là một biện pháp hợp lý để cải thiện tính khách quan của bài kiểm tra. Đồng thời, công tác giảng dạy cần thống nhất giữa các giảng viên để nhấn mạnh sự khác biệt về đặc điểm hình thái nhiễm sắc thể qua các kỳ ở nguyên phân. Tương tự, với 2 câu (95 và 118) ở nội dung liên quan đến trạng thái NST của giám phân cũng cần được xem xét chỉnh sửa hoặc loại bỏ khỏi bài kiểm tra thường xuyên trong ngân hàng câu hỏi.

4. Kết luận

Việc bước đầu áp dụng mô hình 2 tham số IRT vào đánh giá chất lượng câu hỏi MCQ giúp nâng cao tính khách quan và độ chính xác trong đo lường năng lực người học, đảm bảo chất lượng câu hỏi trong ngân hàng. Tuy nhiên, do nghiên cứu được thực hiện trên bài kiểm tra trực tuyến với thời gian làm bài kéo dài cho bài kiểm tra thường xuyên và thiếu kiểm soát chặt chẽ trong điều kiện làm bài online do người học chuẩn bị điều kiện làm bài nên kết quả đánh giá câu hỏi có thể bị ảnh hưởng. Do đó, cần tiếp tục kiểm định chất lượng câu hỏi thông qua bài thi cuối kỳ với điều kiện giám sát nghiêm ngặt hơn. Đồng thời, việc mở rộng nghiên cứu sang nhiều chủ đề khác và chọn mô hình IRT phù hợp cho toàn bộ ngân hàng sẽ góp phần nâng cao độ chính xác và tính toàn diện trong đánh giá năng lực người học.

Lời cảm ơn

Chân thành cảm ơn sinh viên trường Đại học Y Dược Cần Thơ đã hoàn thành bài kiểm tra thường xuyên để có dữ liệu phân tích. Đồng thời, cảm ơn bộ môn Sinh học - Di truyền và Trường Đại học Y Dược Cần Thơ đã tạo điều kiện thuận lợi cho nghiên cứu.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] T. B. H. Le, D. D. Hung, T. K. N. Pham, and T. H. M. Bui, "A solution to support the analysis and evaluation of multiple-choice questions on the Moodle system using the IRT model," (in Vietnamese), *Journal of Education*, vol. 22, no. 16, pp. 17-23, 2022.
- [2] W. N. Arifin and M. S. B. Yusoff, "Item response theory for medical educationists," *Education in Medicine Journal*, vol. 9, no. 3, pp. 69-81, 2017, doi: 10.21315/eimj2017.9.3.8.
- [3] V. C. Nguyen and V. T. Pham, "Analysis and evaluation of multiple-choice test items and test design: A study on application of item response theory," *TNU Journal of Science and Technology*, vol. 226, no. 13, pp. 72-81, 2021.
- [4] Q. T. Le, *Tests and Applications*, Hanoi: Science and Technology Publishing House, (in Vietnamese), 2008.
- [5] B. H. T. Nguyen and T. T. Nguyen, "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," (in Vietnamese), *Journal of Science and Technology - The University of Da Nang*, vol. 21, no. 4, pp. 34-39, 2023.
- [6] A. David and M. Ida, *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences*, Springer, 2019.
- [7] H. Khoshsima and S. M. H. Toroujeni, "Computer Adaptive Testing (CAT) Design; Testing Algorithm and Administration Mode Investigation," *European Journal of Education Studies*, vol. 3, no. 5, 2017, doi: 10.5281/zenodo.576047.
- [8] G. A. Fischer and I. W. Molenaar, *Rasch Models: Foundations, Recent Developments, and Applications*, New York: Springer, 1995.
- [9] F. B. Baker, *The Basics of Item Response Theory*, ERIC Clearinghouse on Assessment and Evaluation, 2001.
- [10] J. Mahmud, "Item response theory: A basic concept," *Educational Research and Reviews*, vol. 12, no. 5, pp. 258-266, 2017.
- [11] I. Paek and K. Cole, *Using R for Item Response Theory Model Applications*, New York: Routledge, 2019.
- [12] V. Mutiawani, A. M. Athaya, K. Saputra, and M. Subianto, "Implementing Item Response Theory (IRT) Method in Quiz Assessment System" *TEM Journal*, vol. 11, no. 1, pp. 210-218, 2022.
- [13] O. M. Adetutu and H. B. Lawal, "Applications of Item Response Theory Models to Assess Item Properties and Students' Abilities in Dichotomous Responses Items," *Open Journal of Educational Development*, vol. 3, no. 1, pp. 1-19, 2023, doi: 10.52417/ojed.v3i1.304.
- [14] H. C. Do, A. V. Lu, and H. U. Phan, "Applying 3-parameter logistic model in validating the level of difficulty, discrimination and guessing of items in a multiple choice test," (in Vietnamese), *Journal of Science - Ho Chi Minh City University of Education*, vol. 85, no. 7, pp. 174-184, 2016, doi: 10.54607/hcmue.js.0.7(85).425(2016).

PHỤ LỤC

| Câu hỏi | Độ khó | Độ phân biệt | Nhóm độ khó | Nhóm độ phân biệt |
|----------------|---------------|---------------------|--------------------|--------------------------|
| Item8 | -1,022735693 | 1,486703354 | Dễ | Tốt |
| Item12 | -0,479066276 | 1,354535629 | Trung bình | Tốt |
| Item21 | -1,419228339 | 1,660852217 | Dễ | Tốt |
| Item23 | -1,710779169 | 1,377874512 | Dễ | Tốt |
| Item25 | -1,020252873 | 1,445216553 | Dễ | Tốt |
| Item26 | -1,728093937 | 1,497963512 | Dễ | Tốt |
| Item41 | -1,804321078 | 1,588240663 | Dễ | Tốt |
| Item51 | -1,457803871 | 1,677295248 | Dễ | Tốt |
| Item56 | -1,068285284 | 1,614930186 | Dễ | Tốt |
| Item57 | -0,995446852 | 1,653752864 | Trung bình | Tốt |
| Item58 | -1,174662292 | 1,47251704 | Dễ | Tốt |
| Item59 | -0,785580221 | 1,656546346 | Trung bình | Tốt |
| Item60 | -1,360864944 | 1,454153114 | Dễ | Tốt |
| Item63 | -1,312224441 | 1,613177989 | Dễ | Tốt |
| Item64 | -1,189223024 | 1,591747635 | Dễ | Tốt |
| Item69 | -1,066816621 | 1,646833638 | Dễ | Tốt |

| | | | | |
|---------|--------------|-------------|------------|---------|
| Item70 | -1,365738822 | 1,446623309 | Dễ | Tốt |
| Item72 | -1,23941635 | 1,472117266 | Dễ | Tốt |
| Item74 | -1,61570724 | 1,607674598 | Dễ | Tốt |
| Item78 | -1,348997108 | 1,4962834 | Dễ | Tốt |
| Item79 | -1,059457094 | 1,415031647 | Dễ | Tốt |
| Item80 | -1,420000417 | 1,630686642 | Dễ | Tốt |
| Item81 | -1,146737393 | 1,454671689 | Dễ | Tốt |
| Item83 | -1,378347506 | 1,652591311 | Dễ | Tốt |
| Item85 | -0,778042813 | 1,431918878 | Trung bình | Tốt |
| Item86 | -1,416618214 | 1,530370626 | Dễ | Tốt |
| Item88 | -1,495770381 | 1,636482856 | Dễ | Tốt |
| Item89 | -1,557942491 | 1,675694733 | Dễ | Tốt |
| Item90 | -1,654057407 | 1,548409171 | Dễ | Tốt |
| Item91 | -1,562773127 | 1,667289805 | Dễ | Tốt |
| Item93 | -1,455238041 | 1,381722689 | Dễ | Tốt |
| Item99 | -1,685547386 | 1,503255361 | Dễ | Tốt |
| Item100 | -1,49226875 | 1,508722718 | Dễ | Tốt |
| Item102 | -1,456832117 | 1,592755094 | Dễ | Tốt |
| Item104 | -1,825096857 | 1,68783299 | Dễ | Tốt |
| Item106 | -1,49745305 | 1,52538831 | Dễ | Tốt |
| Item109 | -0,698910947 | 1,626579312 | Trung bình | Tốt |
| Item115 | -1,369989612 | 1,559190438 | Dễ | Tốt |
| Item117 | -1,188810292 | 1,447173481 | Dễ | Tốt |
| Item1 | -1,65998 | 1,873300445 | Dễ | Rất tốt |
| Item3 | -1,92942835 | 1,935329915 | Dễ | Rất tốt |
| Item4 | -1,865980995 | 2,136809013 | Dễ | Rất tốt |
| Item6 | -1,717107088 | 1,800017258 | Dễ | Rất tốt |
| Item10 | -1,116015004 | 1,730879111 | Dễ | Rất tốt |
| Item14 | -1,419670702 | 1,937922129 | Dễ | Rất tốt |
| Item16 | -2,269673016 | 2,061730418 | Rất dễ | Rất tốt |
| Item17 | -1,407666016 | 1,779744309 | Dễ | Rất tốt |
| Item18 | -1,707963081 | 1,741502253 | Dễ | Rất tốt |
| Item19 | -1,249009759 | 2,091889116 | Dễ | Rất tốt |
| Item22 | -1,16031324 | 1,743430724 | Dễ | Rất tốt |
| Item24 | -1,323861133 | 2,178739701 | Dễ | Rất tốt |
| Item28 | -1,442568686 | 2,721969893 | Dễ | Rất tốt |
| Item29 | -1,844807621 | 1,808917305 | Dễ | Rất tốt |
| Item30 | -1,560462832 | 2,065732688 | Dễ | Rất tốt |
| Item31 | -1,450986657 | 2,17850204 | Dễ | Rất tốt |
| Item32 | -1,158529662 | 1,958828455 | Dễ | Rất tốt |
| Item33 | -1,662911344 | 1,716411085 | Dễ | Rất tốt |
| Item39 | -1,437450629 | 1,893131078 | Dễ | Rất tốt |
| Item42 | -1,532029287 | 2,401876562 | Dễ | Rất tốt |
| Item43 | -1,69794251 | 1,837006093 | Dễ | Rất tốt |
| Item44 | -1,540968841 | 2,299257015 | Dễ | Rất tốt |
| Item45 | -1,548664224 | 2,580460264 | Dễ | Rất tốt |
| Item46 | -1,185718788 | 2,456557209 | Dễ | Rất tốt |
| Item47 | -1,499823046 | 2,312117534 | Dễ | Rất tốt |
| Item48 | -1,474949475 | 2,212792602 | Dễ | Rất tốt |
| Item49 | -1,340146702 | 2,917149856 | Dễ | Rất tốt |
| Item50 | -1,727538402 | 1,861207531 | Dễ | Rất tốt |
| Item52 | -1,527822086 | 2,159203326 | Dễ | Rất tốt |

| | | | | |
|----------------|---------------------|---------------------|-------------------|----------------|
| Item54 | -1,499760239 | 2,189643279 | Dễ | Rất tốt |
| Item55 | -1,215356754 | 1,798510384 | Dễ | Rất tốt |
| Item61 | -1,108801196 | 1,718688062 | Dễ | Rất tốt |
| Item62 | -1,50556174 | 2,018357412 | Dễ | Rất tốt |
| Item65 | -1,307195042 | 1,740789314 | Dễ | Rất tốt |
| Item66 | -1,426704456 | 1,706630799 | Dễ | Rất tốt |
| Item67 | -1,38942059 | 2,473525548 | Dễ | Rất tốt |
| Item73 | -1,277918466 | 1,809931641 | Dễ | Rất tốt |
| Item76 | -1,217735055 | 1,897634844 | Dễ | Rất tốt |
| Item77 | -1,614271123 | 1,84728517 | Dễ | Rất tốt |
| Item82 | -1,450102958 | 2,181481507 | Dễ | Rất tốt |
| Item84 | -1,513125263 | 2,148498182 | Dễ | Rất tốt |
| Item87 | -1,682696868 | 2,311765927 | Dễ | Rất tốt |
| Item92 | -1,576257935 | 1,774505867 | Dễ | Rất tốt |
| Item98 | -1,576871015 | 2,125506515 | Dễ | Rất tốt |
| Item103 | -1,782730517 | 1,758115112 | Dễ | Rất tốt |
| Item107 | -1,182454716 | 2,137702878 | Dễ | Rất tốt |
| Item108 | -1,149717512 | 1,986879565 | Dễ | Rất tốt |
| Item110 | -0,80663088 | 1,71416787 | Trung bình | Rất tốt |
| Item111 | -1,251454765 | 1,742766001 | Dễ | Rất tốt |
| Item116 | -1,078610022 | 1,941459932 | Dễ | Rất tốt |
| Item15 | 0,334648616 | -0,454787786 | Trung bình | Rất kém |
| Item27 | 8,585575151 | 0,041380952 | Rất khó | Rất kém |
| Item68 | -2,829651421 | 0,327057014 | Rất dễ | Rất kém |
| Item34 | -3,018889087 | 0,540538337 | Rất dễ | Kém |
| Item95 | 0,772993336 | 0,388468029 | Trung bình | Kém |
| Item118 | -1,342571099 | 0,438649787 | Dễ | Kém |
| Item2 | -1,474237354 | 1,132914587 | Dễ | Bình thường |
| Item5 | -1,535154388 | 1,246881364 | Dễ | Bình thường |
| Item7 | -1,140334282 | 1,083994185 | Dễ | Bình thường |
| Item9 | -0,805877978 | 1,240296748 | Trung bình | Bình thường |
| Item11 | -1,540232564 | 0,783609674 | Dễ | Bình thường |
| Item13 | -1,89498276 | 0,823094618 | Dễ | Bình thường |
| Item20 | -2,101505178 | 0,901687707 | Rất dễ | Bình thường |
| Item35 | -0,88543759 | 1,062399752 | Trung bình | Bình thường |
| Item36 | -1,036764095 | 0,985262103 | Dễ | Bình thường |
| Item37 | -0,822482414 | 0,989660604 | Trung bình | Bình thường |
| Item38 | -1,667572303 | 1,343803574 | Dễ | Bình thường |
| Item40 | -1,359001583 | 1,293813512 | Dễ | Bình thường |
| Item53 | -1,352590067 | 1,135781312 | Dễ | Bình thường |
| Item71 | -1,27198987 | 1,280312018 | Dễ | Bình thường |
| Item75 | -0,917514488 | 1,21997787 | Trung bình | Bình thường |
| Item94 | -0,568625164 | 1,109050596 | Trung bình | Bình thường |
| Item96 | -0,255956297 | 0,782130252 | Trung bình | Bình thường |
| Item97 | -1,502491328 | 1,320941533 | Dễ | Bình thường |
| Item101 | -1,491693916 | 0,764413276 | Dễ | Bình thường |
| Item105 | -1,869576373 | 1,063022417 | Dễ | Bình thường |
| Item112 | -1,162630739 | 0,783126081 | Dễ | Bình thường |
| Item113 | -2,618288044 | 1,155231639 | Rất dễ | Bình thường |
| Item114 | -0,39146137 | 1,018302904 | Trung bình | Bình thường |
| Item119 | -1,297721113 | 1,020983455 | Dễ | Bình thường |