

MỐI QUAN HỆ GIỮA ĐỘ KHÓ NĂNG LỰC VÀ ĐỘ PHÂN CÁCH CỦA CÂU HỎI TRẮC NGHIỆM SINH LÝ HỌC TRONG CÁC ĐỀ THI TUYỂN SINH SAU ĐẠI HỌC DÀNH CHO ĐỐI TƯỢNG CHUYÊN KHOA CẤP I TỪ NĂM 2018 - 2022 TẠI ĐẠI HỌC Y DƯỢC THÀNH PHỐ HỒ CHÍ MINH

Phạm Dương Uyển Bình¹, Trần Thị Diệu², Võ Đăng Khoa³, Bùi Thanh Tú³, Nguyễn Đình Chương⁴,
Mai Phương Thảo⁴, Nguyễn Anh Vũ³, Vĩnh Sơn¹, Phạm Lê An⁵

TÓM TẮT

Mục tiêu: Trong giáo dục dựa trên năng lực, việc đánh giá chính xác năng lực của thí sinh là quan trọng vì đánh giá giá trị và tin cậy thúc đẩy các hoạt động học tập và đảm bảo năng lực trong các lĩnh vực công việc nhằm cung cấp cho cộng đồng nguồn nhân lực đáng tin cậy. Giả định truyền thống cho rằng các câu hỏi dễ trong bài kiểm tra có độ phân cách thấp, hạn chế trong việc phân biệt giữa học viên ở các cấp độ thành thạo khác nhau. Tuy nhiên, có những bằng chứng cho thấy khả năng phân cách của câu hỏi có thể bị ảnh hưởng bởi các yếu tố khác như đặc điểm tiềm tàng của thí sinh như: môi trường học tập, thành tích, v.v... Do đó, nghiên cứu này phân tích câu hỏi trắc nghiệm của bài thi tuyển sinh sau đại học về độ khó và phân cách nhằm mục đích phân tích mối liên hệ phức tạp của độ khó và độ phân cách cũng như đặc điểm tiềm tàng của thí sinh trong ngữ cảnh của đề thi trắc nghiệm đầu vào Chuyên khoa cấp I môn Sinh lý học.

Đối tượng và phương pháp nghiên cứu: Nghiên cứu cắt ngang được áp dụng để phân tích toàn diện 5 bài kiểm tra Sinh lý học mỗi bài gồm 120 câu, được thực hiện trong vòng 5 năm, từ năm 2018 đến 2022. Nghiên cứu đã thu thập 3928 kết quả kiểm tra từ một nhóm đa dạng thí sinh dự thi sau đại học. Phân tích câu đã được thực hiện để xác định độ khó theo CTT và IRT cùng chỉ số phân cách r_{pbis} trên phần mềm phân tích Basicstat. Ngoài ra, các phép kiểm Anova dùng để so sánh số câu trả lời đúng giữa các đặc điểm mẫu khác nhau như trường, xếp loại tốt nghiệp và năm tốt nghiệp với phần mềm SPSS 20.

Kết quả: Phần lớn trong số 600 câu của bài thi môn Sinh lý học được phân loại là dễ hoặc rất dễ. Tỷ lệ câu hỏi trong vòng 5 năm có chỉ số phân cách r_{pbis} tốt trên 0,3 thuộc số câu dễ-rất dễ theo độ khó logit Rasch: 30,8% đến 43,3% và theo độ khó CTT 52,5 đến 71,1 %. Kết quả phân tích đặc điểm của mẫu và số câu làm đúng cho thấy sự khác biệt có ý nghĩa về số câu trả lời đúng giữa các thí sinh tốt nghiệp từ các trường đại học khác nhau [$F(5,3921) = 36,694, p < 0,001$], xếp loại tốt nghiệp [$F(5, 3921) = 241,237, p < 0,001$] và năm tốt nghiệp khác nhau [$F(3,3923) = 36,082, p < 0,001$].

Kết luận: Tỷ lệ khá cao các câu hỏi MCQ dễ trong 5 bài kiểm tra có độ phân cách tốt. Điều này làm nổi bật mối quan hệ phức tạp giữa độ khó và độ phân cách câu. Sự khác biệt về trường đã học, học lực, v.v... đôi với điểm số minh chứng cho tồn tại đặc điểm tiềm tàng (năng lực) của thí sinh.

Từ khóa: độ khó, logit, độ phân cách, r_{pbis} , CTT, IRT, Rasch

¹Phòng Đảm bảo chất lượng, Đại học Y Dược TP. Hồ Chí Minh

²Bộ môn Tin học, Khoa Khoa học cơ bản, Đại học Y Dược TP. Hồ Chí Minh

⁴Phòng Đào tạo Sau Đại học, Đại học Y Dược TP. Hồ Chí Minh

³Bộ môn Toán, Khoa Khoa học cơ bản, Đại học Y Dược TP. Hồ Chí Minh

⁵Trung tâm Y học Gia đình, Đại học Y Dược TP. Hồ Chí Minh

Tác giả liên lạc: TS. Phạm Dương Uyển Bình ĐT: 0961156505 Email: uyenbinh@ump.edu.vn

Tạp chí Y học Thành phố Hồ Chí Minh, 27(1):170-176. DOI: [10.32895/hcjm.m.2024.01.24](https://doi.org/10.32895/hcjm.m.2024.01.24)

ABSTRACT

EXPLORING THE COMPLEX RELATIONSHIP BETWEEN ITEM DIFFICULTY AND DISCRIMINATION INDEX IN PHYSIOLOGY MCQS IN THE POST-GRADUATE ADMISSION TEST FOR 1ST SPECIALIST FROM 2018 TO 2022 AT UNIVERSITY OF MEDICINE AND PHARMACY IN HO CHI MINH

Pham Duong Uyen Binh, Tran Thi Dieu, Vo Dang Khoa, Bui Anh Tu, Nguyen Dinh Chuong, Mai Phuong Thao, Nguyen Anh Vu, Vinh Son, Pham Le An

* Ho Chi Minh City Journal of Medicine * Vol. 27 - No. 1 - 2024: 170 - 176

Objectives: In competency-based education, accurate assessment of candidates' abilities is important because valid and reliable assessments promote learning activities and ensure competency in specific areas of work to reliably provide the community with human resources. The traditional assumption is that easy questions on a test will have low separation, limited in their ability to differentiate between students at different proficiency levels. However, there is evidence that the ability of questions to separate can be influenced by factors other than question difficulty such as candidate's potential characteristics such as learning environment, achievement, etc. Therefore, this study analyzes the multiple-choice questions of the graduate admission test in terms of difficulty and separation with the aim of analyzing the complex relationship of difficulty and separation as well as the underlying characteristics of candidates in the context of the Physiology exam to enroll in the Level I specialist doctor training program.

Methods: A cross-sectional study was done to thoroughly assess 5 Physiology examinations, each consisting of 120 questions. The tests were administered over a period of 5 years, from 2018 to 2022. The study gathered 3928 answers sheet from a varied cohort of graduate examinees. Analyzed sentences were evaluated for difficulty using both Classical Test Theory (CTT) and Item Response Theory (IRT), as well as the rpbis discriminant index on the Basicstat analysis software. Furthermore, Anova tests were employed to assess the quantity of accurate responses among several sample attributes, including as school, graduation classification, and year of graduation, utilizing SPSS 20 program.

Results: The majority of the 600 questions on the Physiology exam were classified as easy or very easy. Proportion of questions within 5 years with good rpbis discriminant index above 0.3 among easy -very easy questions according to Rasch logit difficulty: 30.8% to 43.3% and according to CTT difficulty 52.5 to 71.1%. The results of analyzing the characteristics of the sample and the number of correct answers show a significant difference in the number of correct answers between candidates graduating from different universities [$F(5,3921) = 36.694, p < 0.001$], graduation rank [$F(5, 3921) = 241.237, p < 0.001$], and graduation year differed [$F(3,3923) = 36.082, p < 0.001$].

Conclusions: A significant proportion of the multiple-choice questions (MCQ) in the 5 physiological tests in 5 years have good rpbis discriminant index, which highlights the complex relationship between difficulty and sentence discriminant. The difference in school attended, academic performance, etc. in the test sum scores proves the existence of latent trait (abilities) of candidates.

Keywords: item difficulty, item 27% discrimination, r_{pbis} , reliability, Classical Test Theory, Items Response Theory, Rasch

ĐẶT VẤN ĐỀ

Môn Sinh lý học thuộc khối kiến thức y học cơ sở, có vai trò quan trọng đầu vào đối với người học thuộc khối ngành khoa học sức khỏe,

giải thích hoạt động chức năng của cơ thể ở trạng thái bình thường. Môn Sinh lý học được giảng dạy trong giai đoạn tiền lâm sàng; và được vận dụng để giải thích cơ chế bệnh lý, triệu

chứng bệnh, làm tiền đề cho việc chẩn đoán và điều trị trong đào tạo những năm học lâm sàng về sau. Vì vậy, bài thi môn Sinh lý học được lựa chọn để làm môn thi tuyển sinh đánh giá năng lực đầu vào của người dự tuyển vào các bậc đào tạo sau đại học trong các trường Y sinh.

Trong giáo dục dựa trên năng lực, đánh giá năng lực học viên đầu vào được chú trọng nhiều hơn vì đánh giá hợp lệ và đáng tin cậy thúc đẩy các hoạt động học tập và đảm bảo năng lực hoàn thành công việc trong tương lai. Các bài kiểm tra thành quả học tập MCQ là công cụ đo lường thành quả học tập của người học và mức độ năng lực nhận thức của họ trong các môn học khác nhau trong chương trình đào tạo Y sinh. Khi thiết kế các bài kiểm tra như vậy quan trọng là cân nhắc độ khó và độ phân cách của các câu hỏi, vì chúng ảnh hưởng trực tiếp đến giá trị và độ tin cậy của bài kiểm tra đánh giá. Theo quan niệm đo lường truyền thống có mối quan hệ giữa độ khó và chỉ số phân cách, câu MCQ lý tưởng thường có khả năng phân cách cao cũng như độ khó trung bình, các câu MCQ quá dễ hay quá khó trong bài kiểm tra là dấu hiệu của phân cách thấp, không phân biệt đầy đủ giữa các thí sinh có trình độ khác nhau cần cải thiện⁽¹⁾. Tuy nhiên, các nghiên cứu gần đây bắt đầu chứng minh rằng mối quan hệ giữa độ khó của câu hỏi và chỉ số phân cách phức tạp hơn⁽²⁾. Nói cách khác giả định này có thể không đúng đối với tất cả các trường hợp cho thấy các câu hỏi ở mức độ khó và hiểu mặc dù được đánh giá “dễ” hơn so với câu hỏi ở mức độ áp dụng, nhưng lại có thể phân biệt tốt hơn giữa những người học có tổng điểm cao và điểm thấp⁽³⁾.

Các tham số câu hỏi MCQ bao gồm độ khó theo mô hình CTT và IRT, cùng độ phân cách.

Theo lý thuyết trắc nghiệm cổ điển CTT, độ khó là tỷ lệ người làm đúng câu trắc nghiệm chia cho tổng số người tham gia làm bài trắc nghiệm⁽⁴⁾. Độ phân cách D 27% cho thầy cô dạy lớp đánh giá lựa chọn câu cho ngân hàng câu do đơn giản có thể tính với máy tính cầm tay và r_{pbis} cho các cấp độ cao hơn như khoa trường

quốc gia do tính chính xác hơn⁽⁵⁾.

Theo lý thuyết đáp ứng câu hỏi (Items Response Theory) độ khó của câu hỏi là một tham số thể hiện mức độ khó của một câu hỏi đối với người làm bài, thường được tính bằng điểm trên thang đo năng lực logit mà người làm bài có 50% khả năng trả lời đúng câu hỏi. Các mô hình IRT giả định rằng khả năng người làm bài trả lời đúng câu hỏi tăng lên khi năng lực của họ tăng lên⁽⁶⁾. Ví dụ, một câu hỏi có tham số độ khó cao là một câu hỏi mà chỉ những người làm bài có năng lực cao mới có khả năng trả lời đúng. Một câu hỏi có tham số độ khó thấp là một câu hỏi mà ngay cả những người làm bài có năng lực thấp cũng có khả năng trả lời đúng. Độ khó của câu hỏi là một tham số quan trọng trong IRT vì nó cho phép tạo ra các bài kiểm tra được điều chỉnh theo mức độ năng lực của người làm bài. Lý tưởng nhất là, nếu hệ thống phân cấp câu hỏi đại diện cho một phạm vi năng lực đầy đủ và các câu hỏi thường sẽ dao động từ -2 đến +2 logit với giá trị logit càng thấp thì câu hỏi càng dễ⁽⁶⁾. Bên cạnh đó, chỉ số phân cách (r_{pbis}) trên 0,3 cho thấy sự phân biệt hiệu quả giữa các ứng viên mạnh và yếu hơn, góp phần vào khả năng đánh giá năng lực tổng thể của thí sinh thông qua kỳ thi.

Ngoài ra còn có thể có những yếu tố khác góp phần vào khả năng phân biệt năng lực người học, yếu tố đặc điểm tiềm tàng của thí sinh ít được tìm hiểu trong các nghiên cứu đánh giá chất lượng bài kiểm tra. Đặc điểm tiềm tàng là những khả năng, phẩm chất, hay năng lực ẩn sâu trong mỗi con người, tuy không thể nhìn thấy bằng mắt thường nhưng lại đóng vai trò quan trọng trong việc định hình tương lai của mỗi người⁽⁷⁾. Các điểm tiềm tàng có thể là khả năng học tập, thời điểm tốt nghiệp và môi trường học tập⁽⁸⁾. Trong đo lường và đánh giá, đặc điểm tiềm tàng được hé lộ thông qua các bài kiểm tra và bài tập được thiết kế cẩn thận để đánh giá năng lực của mỗi cá nhân để hoàn thành công việc được thiết kế trong chương

trình huấn luyện.

Do đó, nghiên cứu của chúng tôi nhằm:

Khám phá mối quan hệ giữa độ khó và chỉ số phân cách của câu hỏi trong bộ đề thi trắc nghiệm đầu vào Chuyên khoa cấp I môn Sinh lý học và tìm hiểu các đặc điểm tiềm tàng của thí sinh ảnh hưởng đến khả năng phân cách của câu trắc nghiệm.

ĐỐI TƯỢNG-PHƯƠNG PHÁP NGHIÊN CỨU

Đối tượng nghiên cứu

Sinh viên tốt nghiệp từ nhiều trường đại học y khoa ở Việt Nam. Nghiên cứu được thực hiện trong khoảng thời gian 5 năm, từ năm 2018 đến năm 2022 đã thu thập 3928 bài kiểm tra, và việc phân tích các câu hỏi trắc nghiệm một đáp án với mỗi câu có 4 lựa chọn theo hai mô hình CTT và IRT.

Phương pháp nghiên cứu

Thiết kế nghiên cứu

Một nghiên cứu cắt ngang sử dụng để phân tích 120 câu bài kiểm tra Sinh lý học thi đầu vào Chuyên khoa cấp I trên đối tượng sinh viên tốt nghiệp từ nhiều trường đại học Y khoa ở Việt Nam.

Công cụ

Các câu hỏi trắc nghiệm Sinh lý học được sử dụng trong kỳ thi đầu vào sau đại học trong 5 năm từ 2018 đến 2022 đã được phân tích, mỗi bài gồm 120 câu có một câu đúng và ba lựa chọn. Chỉ số độ khó trong mô hình IRT (trên thang đo logit) và chỉ số phân biệt (r_{pbis}) mô hình CTT, đã được sử dụng để đánh giá các câu hỏi trắc nghiệm với phần mềm BasicStat của Đại học Y Dược Thành phố Hồ Chí Minh, cùng các phần mềm thống kê và phân tích như Microsoft Excel, SPSS 20.

Độ khó của các câu hỏi trắc nghiệm Sinh lý học

Độ khó câu hỏi theo Lý thuyết cổ điển (CTT)

Bảng 1. Đặc điểm dân số nghiên cứu

Năm	2018	2019	2020	2021	2022
Số thí sinh	549 (13,98%)	692 (17,62%)	914 (23,27%)	781 (19,88%)	992 (25,25%)
Tổng số câu	120	120	120	120	120

Độ khó p bằng tỷ số phần trăm thí sinh làm đúng câu hỏi trên tổng số thí sinh tham gia làm câu hỏi đó. Độ khó p cho biết tỷ lệ thí sinh trả lời đúng câu hỏi nhiều hay ít. Giá trị p càng lớn thì câu hỏi càng dễ:

$$p = \text{số thí sinh làm đúng} / \text{số thí sinh làm bài}$$

Độ khó dựa trên thang logit theo Lý thuyết ứng đáp câu hỏi (IRT)

Lý tưởng nhất là, nếu hệ thống phân cấp câu hỏi đại diện cho một phạm vi năng lực đầy đủ và các câu hỏi thường sẽ dao động từ -2 đến +2 logit. Logit là đơn vị đo lường thang đo khoảng dọc theo thang đo khoảng thu được từ phân tích Rasch. Dựa trên thang đo logit, chúng tôi tập trung vào 2 phân loại chỉ số độ khó bao gồm câu hỏi dễ (thang đo logit từ -0,1 đến -1); câu hỏi rất dễ (thang đo logit nhỏ hơn -1).

Độ phân cách r_{pbis} được chọn thay cho D 27% do kỳ thi cấp quốc gia và tính bằng cách so sánh điểm trung bình tổng thể của những người trả lời đúng câu hỏi với điểm trung bình tổng thể của những người trả lời sai $r_{pbis} \geq 0,3$ được cho là có chỉ số phân cách tốt.

Thống kê

Thống kê mô tả đã được sử dụng để phân tích dữ liệu. Chỉ số độ khó theo thang đo logit và chỉ số phân biệt r_{pbis} được tính toán bằng phần mềm BasicStat. Mô tả tỷ lệ phần trăm và Phân tích phương sai một chiều (One-way ANOVA) được thực hiện để đánh giá mối quan hệ giữa xuất thân trường tốt nghiệp đại học, xếp loại tốt nghiệp, năm tốt nghiệp và số câu trả lời đúng.

KẾT QUẢ

Qua 5 năm, đề thi tuyển sinh sau đại học môn Sinh lý học có số câu hỏi ổn định 120 câu với 4 lựa chọn. Số thí sinh dự thi tăng nhiều qua 5 năm từ 549 thí sinh năm 2018 đến 992 thí sinh năm 2022 (Bảng 1).

Năm	2018	2019	2020	2021	2022
Số lựa chọn	4	4	4	4	4
Xếp loại tốt nghiệp bậc đại học					
Giỏi	216 (5.48%)				
Khá	1912 (48.49%)				
Trung bình khá	1692 (42.91%)				
Trung bình	123 (3.12%)				
Năm tốt nghiệp đại học					
2018 - 2022	958 (5.5%)				
2016 - 2017	1293 (48.69%)				
2014 - 2015	909 (43.09%)				
2013 về trước	767 (3.13%)				

Bảng 2. Tỷ lệ câu hỏi dễ và rất dễ có độ phân cách $r_{pbis} \geq 0,3$

Năm	$r_{pbis} \geq 0,3$ (N=120)					
	Độ khó logit Rasch			Độ khó CTT		
	-1~-0,1	<-1	Tỷ lệ (%)	0,6~0,8	$\geq 0,8$	Tỷ lệ (%)
2018	28	09	30,8	19	44	52,5
2019	31	15	38,3	26	44	58,3
2020	29	14	35,8	27	45	60
2021	36	13	40,8	24	62	71,6
2022	33	19	43,3	26	52	65

Bảng 3. Sự khác biệt về số câu đúng với các đặc điểm của thí sinh

	M	SD	Trung bình khác biệt	Khoảng tin cậy	p value
Trường tốt nghiệp đại học					
Trường tốt nghiệp 1	91,80	15,01			
Trường tốt nghiệp 2	86,16	17,97	5,64*	3,25 ~ 8,05	<0,001
Trường tốt nghiệp 3	97,36	13,52	-5,56*	-8,08 ~ -3,04	<0,001
Trường tốt nghiệp 4	92,08	16,05	-0,27	-2,41 ~ 1,86	0,999
Trường tốt nghiệp 5	87,25	17,68	4,55*	1,90 ~ 7,20	<0,001
Trường tốt nghiệp 6	86,99	18,00	4,81*	2,38 ~ 7,25	<0,001
Xếp loại tốt nghiệp bậc đại học					
Giỏi	102,01	10,77			
Khá	95,19	13,89	0,54	0,38 ~ 0,71	<0,001
Trung bình khá	84,69	17,09	1,23	1,06 ~ 1,39	<0,001
Trung bình	73,20	17,75	1,88	1,61 ~ 2,15	<0,001
Năm tốt nghiệp đại học					
2018 - 2022	92,32	15,75			
2016 - 2017	93,22	14,88	-0,89	-2,69 ~ 0,89	0,570
2014 - 2015	88,58	17,41	3,74*	1,79 ~ 5,69	<0,001
2013 về trước	86,38	18,22	5,94*	3,90 ~ 7,98	<0,001

Trong vòng 5 năm, mỗi năm đề thi có 120 câu, kết quả cho thấy một lượng câu hỏi có chỉ số phân biệt r_{pbis} tốt ($\geq 0,3$) nhưng xếp loại dễ và rất dễ theo mô hình CTT, dao động từ 52,5% đến 71,6%. Tương tự trên mô hình IRT độ khó theo thang logit, câu hỏi dễ (độ khó logit thấp từ -1 đến -0,1) và rất dễ (độ khó logit rất thấp dưới -1) dao động từ 30,8% đến 43,3%. Điều này cho thấy

trong các câu có phân cách tốt thì tỷ lệ số câu dễ và rất dễ theo mô hình CTT nhiều hơn theo mô hình IRT (Bảng 2).

Kết quả phân tích cho thấy có sự khác biệt đáng kể giữa các nhóm trường tốt nghiệp đại học [$F(5,3921) = 36,694$, $p < 0,001$]. Kiểm tra Tukey HSD sau kiểm định cho thấy số câu trả lời đúng của thí sinh tốt nghiệp (TN) từ trường đại học

đánh mã số 2, 5 và 6 thấp hơn đáng kể so với TN trường mã số 1 ($p < 0,001$) (Bảng 3).

Kết quả phân tích cũng cho thấy có sự khác biệt đáng kể giữa các nhóm phân loại tốt nghiệp [$F(5, 3921) = 241,237, p < 0,001$]. Kiểm tra Tukey HSD sau kiểm định cho thấy thí sinh sau đại học có phân loại tốt nghiệp càng cao thì có số câu trả lời đúng nhiều hơn đáng kể so với những thí sinh có phân loại tốt nghiệp thấp hơn ($p < 0,001$) (Bảng 3).

Ngoài ra, kết quả phân tích cũng cho thấy có sự khác biệt đáng kể giữa các nhóm năm tốt nghiệp [$F(3,3923) = 36,082, p < 0,001$]. Kiểm tra Tukey HSD sau kiểm định cho thấy thí sinh tốt nghiệp đại học gần đây từ năm 2018 đến năm 2022 có số câu trả lời đúng cao hơn đáng kể so với những thí sinh tốt nghiệp trước năm 2014 ($p < 0,001$) (Bảng 3).

BÀN LUẬN

Nghiên cứu sử dụng phân tích tham số câu MCQ là một công cụ hiệu quả trong việc quyết định số lượng lựa chọn, lưu giữ các câu hỏi và so sánh thành tích của sinh viên⁽⁹⁾. Chúng tôi áp dụng phân tích câu MCQ để khám phá mối quan hệ giữa độ khó của câu hỏi theo mô hình CTT và mô hình IRT và chỉ số phân cách r_{pbis} trong bài thi đầu vào chương trình bác sĩ chuyên khoa cấp I, dựa trên môn thi Sinh lý học.

Trái ngược với giả định rằng các câu MCQ dễ dựa trên giá trị độ khó cao hay số người làm đúng nhiều thì ít có khả năng phân cách giữa những người học đạt điểm cao và thấp do cả hai nhóm đều làm đúng trên công thức tính $D 27\%$. Tuy nhiên nghiên cứu của chúng tôi cho thấy rằng câu dễ sử dụng độ khó trên CTT cũng như độ khó logit trên IRT vẫn có tiềm năng phân cách năng lực thí sinh khi sử dụng chỉ số phân cách r_{pbis} do tính toán số liệu toàn bộ 100%. Ngoài ra, những phát hiện cho thấy rằng các yếu tố tiềm tàng khác ngoài độ khó, chẳng hạn như cơ sở đào tạo ở bậc đại học, xếp loại tốt nghiệp đại học, thời gian tốt nghiệp đại học đến khi đăng ký dự thi bậc đào tạo sau đại học, có thể

đóng vai trò quan trọng trong việc ảnh hưởng đến độ phân cách. Điều này nhấn mạnh tầm quan trọng của việc xem xét đa yếu tố khi đánh giá khả năng phân cách của các câu hỏi trắc nghiệm. Nghiên cứu của chúng tôi góp phần bổ sung minh chứng cho thấy độ khó và khả năng phân biệt của câu hỏi không nên được coi là các đặc điểm loại trừ lẫn nhau nói chung, và trong đề kiểm tra môn Sinh lý học nói riêng.

Một số đặc điểm tiềm tàng của thí sinh (năng lực) có thể góp phần tạo ra hiện tượng các câu hỏi được phân loại là dễ nhưng có khả năng phân cách cao.

Cụ thể, chúng tôi thấy sự chênh lệch đáng kể về kết quả thi của thí sinh ở các xếp hạng học lực, năm tốt nghiệp và trường tốt nghiệp đại học khác nhau. Có thể nghĩ đến một trong những nguyên nhân là do kiến thức và kỹ năng của thí sinh đạt xếp loại tốt nghiệp càng tốt thì sẽ có năng lực càng cao hơn và số câu làm đúng của cũng sẽ nhiều hơn. Bên cạnh đó, chúng tôi cũng nghĩ đến những trường hợp thí sinh tốt nghiệp những năm trước đây, có thể có ít nguồn học liệu ôn tập hơn nên kết quả thi không tốt bằng so với các thí sinh mới tốt nghiệp gần đây. Ngoài ra, yếu tố trường tốt nghiệp đại học cũng góp phần không nhỏ tạo nên sự khác biệt năng lực của thí sinh. Sự chênh lệch này có khả năng làm cho sức mạnh phân biệt của các câu hỏi trắc nghiệm (MCQs) trở nên cao hơn, ngay cả khi các câu hỏi dễ. Hiện tượng này xảy ra bởi vì khi có sự chênh lệch đáng kể về năng lực giữa các thí sinh, các câu hỏi có thể phân biệt hiệu quả giữa những thí sinh có năng lực trên trung bình và kém dưới ngưỡng. Trong trường hợp như vậy, ngay cả những câu hỏi dễ cũng có thể có độ phân cách tốt^(10,11,12).

Như vậy, các câu hỏi dễ có thể có khả năng phân cách cao là do chênh lệch trình độ của thí sinh. Điều quan trọng là phải nhận ra rằng chỉ riêng mức độ khó thì không xác định khả năng của một câu hỏi trong việc phân biệt giữa các năng lực. do đó, chúng tôi cần nghiên cứu thêm với phân tích câu theo lý thuyết ứng đáp câu hỏi

IRT, trong đó độ khó câu hỏi được đo lường trên thang logit. Điều này có thể giúp thông tin thêm cho việc phát triển các chiến lược đánh giá đo lường chính xác năng lực của người học và thúc đẩy việc đánh giá công bằng.

KẾT LUẬN

Nghiên cứu của chúng tôi ghi nhận tỷ lệ 1/3 đến 1/2 các câu hỏi dễ trong bài kiểm tra môn Sinh lý học trong 5 năm liên tiếp có độ phân cách tốt. Điều này làm nổi bật mối quan hệ phức tạp giữa độ khó và độ phân cách câu. Sự khác biệt về trường đã học, học lực đối với điểm số minh chứng cho tồn tại đặc điểm tiềm tàng (năng lực) của thí sinh vốn được làm cơ sở cho mô hình IRT.

TÀI LIỆU THAM KHẢO

- De C, José F (2014). Classical test theory and item response theory: Two understandings of one high-stakes performance exam. *Colombian Applied Linguistics Journal*, 16(2):167-184.
- Azzopardi M & Azzopardi C (2019). Relationship between item difficulty level and item discrimination in biology final examinations. *Education and New Developments*.
- Karadag N & Sahin MD (2016). Analysis of the Difficulty and Discrimination Indices of Multiple-Choice Questions According to Cognitive Levels in an Open and Distance Learning Context. *Turkish Online Journal of Educational Technology-TOJET*, 15(4):16-24.
- Musa A, Shaheen S, Elmardi A & Ahmed A (2018). Item difficulty & item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine Khartoum University. *Khartoum Medical Journal*, DOI:10.53332/kmj.v11i2.610.
- Dương Thiệu Tống (1995). Phân tích câu trắc nghiệm: Độ phân cách và độ khó. Trong: Dương Thiệu Tống. Trắc nghiệm và đo lường thành quả học tập, ấn bản lần thứ nhất, trang 121 - 143. *Đại học Tổng hợp*, TP. Hồ Chí Minh.
- Cai L, Choi K, Hansen M & Harrell L (2016). Item response theory. *Annual Review of Statistics and Its Application*, 3:297-321.
- Samejima F (1988). Comprehensive latent trait theory. *Behaviormetrika*, 15:1-24.
- Steyer R, Schmitt M & Eid M (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, 13(5):389-408.
- Amini N, Michoux N, Warnier L, Malcourant E, Coche E & Vande BB (2020). Inclusion of MCQs written by radiology residents in their annual evaluation: innovative method to enhance resident's empowerment? *Insights into Imaging*, 11:1-8.
- Pimentel JL & Villaruz MLA (2020). Comparison of Item Difficulty Estimates in a Basic Statistics Test using Irtm and CTT Software Packages in R. *International Journal of Advanced Computer Science and Applications*, DOI: 10.14569/IJACSA.2020.0110346.
- Fatima S, Ahmed ST, Hashmat S, Iftikhar A & Hashmi TA (2021). The relationship of difficulty and discrimination indices with physiology teaching and learning. *Pak J Med Dent*, 10(2):109-13.
- Barker BA, Donovan NJ, Schubert AD & Walker EA (2017). Using Rasch analysis to examine the item-level psychometrics of the Infant-Toddler Meaningful Auditory Integration Scales. *Speech, Language and Hearing*, 20(3):130-143.

Ngày nhận bài:	13/05/2024
Ngày chấp nhận đăng bài:	10/06/2024
Ngày đăng bài:	11/06/2024