

KHẢ NĂNG SUY LUẬN THỐNG KÊ CỦA HỌC SINH TRUNG HỌC PHỔ THÔNG: PHÂN TÍCH BẰNG MÔ HÌNH RASCH

Phạm Thế Quân^{1,+},
Trần Trung²

¹Trường Đại học Sư phạm Hà Nội 2; ²Học viện Dân tộc
+ Tác giả liên hệ • Email: phamthequan@hpu2.edu.vn

Article history

Received: 10/6/2024

Accepted: 31/7/2024

Published: 05/10/2024

Keywords

Statistical reasoning ability,
students, high school,
analysis, Rasch model

ABSTRACT

Among many models of the Item Response Theory (IRT), the Rasch model is the simplest and most commonly used. Mathematician George Rasch has built a simple question characteristic function model that is widely used today. In this paper, we apply the Rasch model of the Item Response Theory (IRT) to analyze test questions and statistical reasoning ability of high school students. The results of the experimental study were analyzed with quantitative methods, the data were processed with QUEST and SPSS 20.0 software. The research subjects were 108 grade 10 students of Binh Giang High School, Binh Giang District, Hai Duong Province. The research results showed that the test questions had high reliability, the distribution of difficulty of the questions in the test covered the students' abilities. In terms of statistical reasoning, the surveyed students performed relatively well in basic skills such as calculating, reading and understanding charts presenting data. However, there were still limitations in applying statistical knowledge to analyze, explain and draw meaningful conclusions from given data. Another notable point is that there was no significant difference in statistical reasoning ability between male and female students.

1. Mở đầu

Ngày nay, Thống kê và Xác suất (TK&XS) đóng vai trò quan trọng trong mọi lĩnh vực, từ khoa học tự nhiên, kỹ thuật, kinh tế cho đến khoa học xã hội. TK&XS là công cụ giúp con người thu thập, phân tích và giải thích dữ liệu một cách hiệu quả. Khi TK&XS mới được đưa vào Chương trình giáo dục phổ thông 2018, quá trình dạy học chủ yếu tập trung vào thực hiện các phép tính để tìm ra các con số thống kê. Tuy nhiên, nhờ sự phát triển của công nghệ, việc thực hiện các phép tính phức tạp không còn là mục tiêu chính. Thay vào đó, trọng tâm đã dịch chuyển sang việc phát triển tư duy, rèn luyện khả năng suy luận theo hướng thống kê. Ở Việt Nam, có thể kể đến một số nghiên cứu theo xu hướng này như: Trần Đức Chiên (2007) nghiên cứu về rèn luyện tư duy thống kê cho HS; Hoàng Nam Hải (2013) nghiên cứu phát triển năng lực suy luận thống kê (SLTK) cho sinh viên cao đẳng; Quách Thị Sen (2020) quan tâm đến phát triển tư duy thống kê cho sinh viên đại học ngành Dược; Hoàng Lê Minh (2023) với nghiên cứu xây dựng môi trường học tập nội dung thống kê ở trường THPT theo hướng rèn luyện kỹ năng SLTK cho HS,...

Việc đánh giá khả năng SLTK của HS trong dạy học TK&XS cũng được nhiều nhà nghiên cứu quan tâm. Tương tự như trong lĩnh vực giáo dục toán học, việc đánh giá trong dạy học TK&XS nhằm hướng tới các mục tiêu: Theo dõi sự tiến bộ của HS, cải thiện phương pháp dạy học, tổ chức lại quá trình học tập của HS và điều chỉnh nội dung chương trình giảng dạy cho phù hợp (Garfield & Chance, 2000; Garfield, 2003). Đánh giá khả năng SLTK của HS trong dạy học TK&XS đã thay đổi theo định hướng mới, tập trung vào 3 yếu tố chính là hiểu biết thống kê, SLTK và tư duy thống kê (Garfield & Ben-Zvi, 2008).

Trong bài báo này, chúng tôi sử dụng Lí thuyết ứng đáp câu hỏi hay còn gọi là “Lí thuyết hồi đáp” (Item Response Theory - IRT) của Georg Rasch để phân tích đề kiểm tra và khả năng SLTK của HS THPT thông qua 20 câu hỏi trắc nghiệm đã được xây dựng. Các câu hỏi nghiên cứu đặt ra là: (1) Đề kiểm tra được xây dựng có phù hợp để đánh giá khả năng SLTK của HS hay không?; (2) Thực trạng khả năng SLTK của HS THPT như thế nào?; (3) Có sự khác biệt đáng kể nào về khả năng SLTK giữa nhóm HS nam và HS nữ hay không?

2. Kết quả nghiên cứu

2.1. Khả năng suy luận thống kê

SLTK có thể được định nghĩa như là cách con người suy luận với các ý tưởng thống kê và làm cho thông tin thống kê trở nên có ý nghĩa. Điều này liên quan đến việc đưa ra các lí giải dựa trên các tập dữ liệu, biểu diễn hay các

số đặc trưng của dữ liệu. SLTK có thể bao gồm việc kết nối một khái niệm với một khái niệm khác (chẳng hạn như: tâm, độ phân tán), hay có thể kết hợp các ý tưởng về dữ liệu và cơ hội (Garfield et al., 2008). Garfield và Chance (2000) đã đưa ra một số gợi ý để thực hiện đánh giá khả năng SLTK của HS, bao gồm: Nghiên cứu tình huống hoặc bài tập thực tế; Sử dụng sơ đồ tư duy; Phê bình các ý tưởng thống kê hoặc các vấn đề thống kê trong tin tức; Viết báo cáo ngắn gọn; Trả lời câu hỏi trắc nghiệm.

Mặc dù khả năng SLTK có thể được đánh giá tốt nhất thông qua giao tiếp trực tiếp với HS (như phỏng vấn, quan sát), hoặc bằng cách xem xét một hồ sơ học tập chi tiết, sâu rộng (như dự án thống kê), nhưng các bài kiểm tra được thiết kế kỹ càng cũng có thể được sử dụng để thu thập thông tin về khả năng SLTK của HS. Một số bài đánh giá đã được xây dựng và sử dụng rộng rãi như: Bài kiểm tra khả năng SLTK (Statistical Reasoning Assessment - SRA) gồm 20 câu hỏi trắc nghiệm của Garfield (2003); bài đánh giá toàn diện kết quả trong khóa học thống kê đầu tiên gồm 40 câu hỏi trắc nghiệm của Delmas và cộng sự (2007); bài kiểm tra mục tiêu và kết quả liên quan đến học thống kê gồm 20 câu hỏi trắc nghiệm của Sabbag và Zieffler (2015);... Trong các bài kiểm tra này, SRA là công cụ đầu tiên được phát triển và xác nhận để đánh giá một số khía cạnh về SLTK của HS. Trước khi có SRA, hầu hết các công cụ và câu hỏi được sử dụng đều tập trung vào các tính toán chứ không phải vào khả năng suy nghĩ và suy luận của người học khi đối mặt với các vấn đề TK&XS. Đặc điểm chung của các bài kiểm tra là đều sử dụng hình thức câu hỏi trắc nghiệm nhiều lựa chọn, riêng Garfield (2003) yêu cầu HS phải giải thích cho lựa chọn của mình. Tuy nhiên, đối tượng chính mà các bài kiểm tra SLTK ở trên hướng đến là sinh viên đại học.

2.2. Sơ lược về mô hình Rasch

Lý thuyết trắc nghiệm cổ điển (Classical Test Theory - CTT) ra đời vào khoảng cuối thế kỉ XIX và hoàn thiện vào những năm 60 của thế kỉ XX. Lý thuyết này được xây dựng dựa trên khoa học thống kê và được ứng dụng chủ yếu trong việc phân tích, đánh giá câu hỏi trắc nghiệm khách quan. Việc đánh giá câu hỏi trắc nghiệm khách quan theo CTT chủ yếu dựa trên các tham số độ khó, độ phân biệt và hệ số tương quan của câu hỏi với đề thi sau khi có kết quả phản hồi của thí sinh đối với các câu hỏi. CTT giúp đơn giản hóa việc đo lường và đánh giá đề thi nhờ khả năng vận hành mô hình mà không cần giả thiết phức tạp. Tuy nhiên, hạn chế lớn nhất của CTT là không tách biệt được các đặc trưng của thí sinh độc lập với các đặc trưng của đề thi, đặc trưng này chỉ có thể giải thích trong mối quan hệ với đặc trưng kia (Lâm Quang Thiệp, 2010). Cụ thể, đặc trưng của đề thi (độ khó, độ phân biệt, độ tin cậy, độ giá trị) phụ thuộc vào nhóm thí sinh làm bài. Do đó, khi áp dụng cho nhóm thí sinh khác, các đặc trưng này sẽ thay đổi, ảnh hưởng đến kết quả đánh giá. Năng lực thí sinh cũng biến đổi theo độ khó của đề thi. Ví dụ, đề dễ sẽ khiến thí sinh được đánh giá cao năng lực và ngược lại với đề khó.

Để khắc phục những điểm yếu kể trên của CTT, trong nửa sau của thế kỉ XX, nhiều nhà nghiên cứu về đo lường trong giáo dục đã đề xuất và cho ra đời lý thuyết ứng đáp câu hỏi (Item Response Theory - IRT) (Morales, 2009). Lý thuyết ứng đáp câu hỏi được xây dựng dựa trên hai giả thiết: Sự ứng đáp của một thí sinh đối với một câu hỏi có thể được tiên đoán bằng năng lực tiềm ẩn của thí sinh; Quan hệ giữa sự ứng đáp câu hỏi của thí sinh và năng lực tiềm ẩn làm cơ sở cho sự đáp ứng đó có thể mô tả bằng một hàm đặc trưng câu hỏi đồng biến (Lâm Quang Thiệp, 2010). Lý thuyết ứng đáp câu hỏi IRT có nhiều mô hình, trong đó Rasch là mô hình đơn giản và được sử dụng phổ biến. Nhà toán học George Rasch đã xây dựng được một mô hình hàm đặc trưng câu hỏi đơn giản và được sử dụng nhiều cho đến ngày nay. Để biểu diễn câu hỏi, George Rasch chỉ chọn một tham số là độ khó của câu hỏi. Mô hình Rasch dựa vào giả thuyết sau: *Một người có năng lực cao hơn một người khác thì xác suất để người đó trả lời đúng một câu hỏi bất kì phải lớn hơn xác suất của người sau; cũng tương tự như vậy, một câu hỏi khó hơn một câu hỏi khác có nghĩa là xác suất để một người bất kì trả lời đúng câu hỏi đó phải bé hơn xác suất để trả lời đúng câu hỏi sau* (Georg, 1960).

Từ giả thuyết trên, có thể thấy xác suất để một thí sinh trả lời đúng một câu hỏi nào đó phụ thuộc vào tương quan giữa năng lực của thí sinh và độ khó của câu hỏi. Công thức của mô hình Rasch như sau:

$$P(X_{ij} = 1 | \theta_i, b_j) = \frac{e^{\theta_i - b_j}}{1 + e^{\theta_i - b_j}}; \text{ trong đó } \theta_i \text{ là năng lực của thí sinh } i, b_j \text{ là độ khó của câu hỏi thứ } j, X_{ij} \text{ là trả}$$

lời của thí sinh thứ i đối với câu hỏi thứ j . Nếu thí sinh trả lời đúng câu hỏi thì $X_{ij} = 1$, nếu trả lời sai thì $X_{ij} = 0$.

Trong mô hình Rasch, xác suất trả lời đúng câu hỏi của thí sinh tỉ lệ thuận với năng lực của thí sinh đó. Xác suất này sẽ tiến dần về 1 khi năng lực của thí sinh tiến đến dương vô cùng. Nếu năng lực của thí sinh bằng với độ khó của câu hỏi thì khả năng trả lời đúng câu hỏi của thí sinh đó là 50%. Mức năng lực này được gọi là ngưỡng của câu hỏi. Như vậy, độ khó của mỗi câu hỏi chính là ngưỡng mà với năng lực đó, khả năng trả lời đúng câu hỏi của thí sinh là

50%. Baker (2001) cho rằng, điểm nổi bật của mô hình Rasch là đã mô tả được mối liên hệ giữa năng lực của mỗi thí sinh đối với các tham số đặc trưng của các câu hỏi thông qua sự ứng đáp của mỗi thí sinh khi trả lời các câu hỏi trong đề thi.

2.3. Nghiên cứu thực nghiệm

2.3.1. Đối tượng, mục đích và phương pháp nghiên cứu thực nghiệm

Đối tượng nghiên cứu là 108 HS lớp 10 Trường THPT Bình Giang, tỉnh Hải Dương (xem bảng 1). Thực nghiệm được tiến hành vào tháng 4/2024.

Bảng 1. Thống kê đặc điểm đối tượng nghiên cứu

Đặc điểm	Số lượng	Tỉ lệ (%)
Giới tính	Nam	58
	Nữ	50
Tổng	108	100

Về giới tính, có 58 HS nam (chiếm 53,7%), 50 HS nữ (chiếm 46,3%). Có thể thấy sự phân bố theo giới tính là tương đối đồng đều. Công cụ được sử dụng trong nghiên cứu này là một bài kiểm tra SLTK gồm 20 câu hỏi trắc nghiệm. Bài kiểm tra được xây dựng dựa trên các căn cứ: Các loại SLTK và biểu hiện; Nội dung và yêu cầu cần đạt của mạch kiến thức TK&XS theo Chương trình giáo dục phổ thông 2018; Các bài đánh giá khả năng SLTK của HS. Bài kiểm tra gồm 4 chủ đề với số lượng các câu hỏi cụ thể trong như trong bảng 2:

Bảng 2. Thành phần các câu hỏi trong bài kiểm tra

TT	Chủ đề	Số lượng câu hỏi	Tỉ lệ (%)
1	Thu thập và mô tả dữ liệu	2	10
2	Tổ chức và thu gọn dữ liệu	8	40
3	Biểu diễn dữ liệu	2	10
4	Phân tích, diễn giải dữ liệu và kết luận	8	40

Mục đích của thực nghiệm là đánh giá khả năng SLTK của HS THPT thông qua mô hình Rasch. Nghiên cứu thực nghiệm sử dụng phương pháp đánh giá định lượng, các số liệu được xử lý với phần mềm QUEST và SPSS 20.0.

2.3.2. Kết quả nghiên cứu thực nghiệm

- *Độ tin cậy của đề kiểm tra và mức độ phù hợp với mô hình Rasch*: Kết quả điểm kiểm tra của HS được thống kê trong bảng 3:

Bảng 3. Thống kê kết quả điểm kiểm tra

Số lượng	Hợp lệ	108
	Bò sót	0
Trung bình		5,694
Trung vị		5,500
Độ lệch chuẩn		1,3822
Phần trăm	25	4,500
	50	5,500
	75	6,875

Lý thuyết hồi đáp IRT có nhiều mô hình đo lường khác nhau. Mô hình Rasch chỉ được áp dụng nếu dữ liệu làm bài của HS phù hợp với mô hình này. Để đánh giá sự phù hợp, cần xem xét hai tiêu chí: Giá trị kì vọng của các bình phương trung bình (mean square) gần bằng 1 (giá trị chấp nhận trong khoảng 0,77 đến 1,3) và độ lệch chuẩn (SD) gần bằng 0 (Adams & Khoo, 1996). Các số liệu về giá trị trung bình (mean) và độ lệch chuẩn (SD) có được khi xử lý dữ liệu bằng phần mềm QUEST cho thấy dữ liệu phù hợp với mô hình Rasch (xem hình 1).

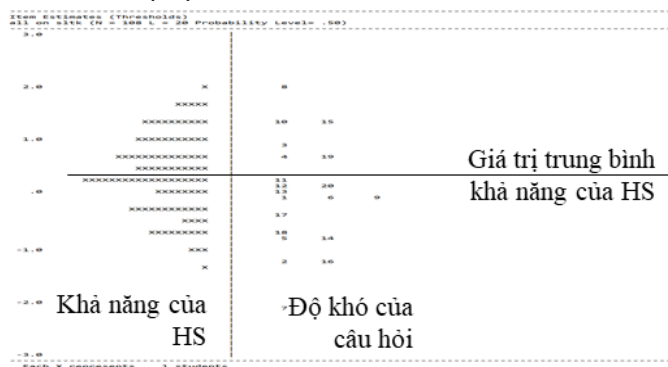
Summary of item Estimates		Summary of case Estimates	
=====		=====	
Mean	.00	Mean	.36
SD	1.02	SD	.72
SD (adjusted)	.99	SD (adjusted)	.50
Reliability of estimate	.95	Reliability of estimate	.48
Fit Statistics		Fit Statistics	
=====		=====	
Infit Mean Square		Infit Mean Square	
Mean	1.00	Mean	1.00
SD	.06	SD	.20
Outfit Mean Square		Outfit Mean Square	
Mean	.98	Mean	.98
SD	.14	SD	.32

Hình 1. Thông tin tổng hợp về đề thi và thí sinh

Độ tin cậy của đề kiểm tra (Reliability of estimate) là 0,95. Do vậy, đây là một đề kiểm tra có độ tin cậy cao và kết quả đo được phản ánh đúng khả năng của HS tham gia làm bài.

- *Mức độ phù hợp của các câu hỏi:* Các câu hỏi phù hợp với mô hình Rasch nếu có chỉ số Infit MNSQ nằm trong khoảng từ 0,77-1,30 (Adams & Khoo, 1996). Trong biểu đồ Item Fit, mỗi câu hỏi được biểu thị bằng một dấu *. Câu hỏi phù hợp sẽ có dấu * tương ứng nằm giữa hai đường chấm thẳng đứng ở vị trí 0,77 và 1,30. Kết được phân tích cho thấy, tất cả các câu hỏi đều phù hợp, không có câu hỏi nào bị loại bỏ.

- *Phân bố khả năng suy luận thống kê của HS và độ khó câu hỏi:* Phần mềm QUEST cung cấp một biểu đồ dưới dạng ma trận (xem biểu đồ 1) cho thấy khả năng của 108 HS và độ khó của 20 câu hỏi. Trong đó, các con số bên phải đường kẻ thẳng của biểu đồ là các câu hỏi, vị trí của số cho biết độ khó của câu hỏi, các câu hỏi ở trên có độ khó cao hơn các câu hỏi ở dưới. Mỗi dấu X nằm bên trái đường kẻ thẳng của biểu đồ đại diện cho một HS, vị trí của dấu X cho biết khả năng SLTK của HS tương ứng. Dấu X ở phía trên đại diện cho HS có khả năng cao hơn.

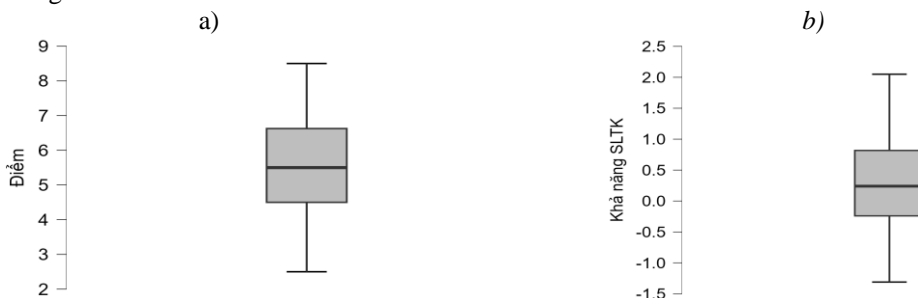


Biểu đồ 1. Biểu đồ phân bố khả năng của HS và độ khó câu hỏi thi

Quan sát phân bố độ khó của câu hỏi và khả năng của HS cho thấy mức độ phù hợp của đề thi đối với những HS dự thi. Một đề thi có chất lượng tốt khi các câu hỏi trong đề thi có độ khó tương ứng với khả năng của HS, kể cả những HS có khả năng thấp và những HS có khả năng cao.

Khả năng của HS và độ khó của câu hỏi được đánh giá bằng thang đo logit, kết quả phân tích thu được khả năng của HS đạt giá trị trong đoạn [-1,31; 2,05], độ khó của câu hỏi đạt giá trị trong đoạn [-2,13; 1,97]. Câu hỏi khó nhất trong đề thi là câu 8, với độ khó 1,97, câu hỏi dễ nhất là câu 7 với độ khó -2,13. Nhìn trên biểu đồ có thể thấy rõ nét sự phân bố về độ khó các câu hỏi thi bao trùm được khả năng của HS. Thông tin tổng hợp về đề thi và thí sinh được phần mềm QUEST tính toán và thu được khả năng trung bình của mẫu HS (case estimate) tham gia bài kiểm tra là 0,36, cao hơn so với độ khó chung của bài kiểm tra nhưng khoảng chênh lệch không nhiều. Nói cách khác, bài kiểm tra tương đối phù hợp với khả năng của HS. Khả năng của HS có cao hơn dù không đáng kể so với độ khó trung bình của các câu hỏi.

Biểu đồ 2 cho thấy, khả năng SLTK của nhóm HS phân bố khá đồng đều. Có 52 HS (chiếm 48%) có khả năng trên mức trung bình.



Biểu đồ 2. Biểu đồ hộp của điểm kiểm tra (a) và biến khả năng SLTK (b)

HS có khả năng SLTK cao nhất (với điểm số là 2,05) là HS nam, số câu trả lời đúng là 17/20. HS này làm sai các câu 3, 17, 19. HS có khả năng SLTK thấp nhất (-1,31) là HS nam, số câu trả lời đúng là 5/20. Câu số 7 được xếp loại rất dễ, mức độ khó thấp hơn nhiều so với khả năng của HS. Hai câu hỏi có độ khó thấp nhất trong đề thi là câu 7 và câu 2 với nội dung như sau:

Câu 7: Tốp 1 gồm 10 HS và điểm số của các HS này trong một bài kiểm tra môn Ngữ văn như sau:

Điểm số	5	6	7	8	9
Số HS	3	2	1	2	2

Điểm kiểm tra môn Ngữ văn trung bình của các HS tốp 1 là bao nhiêu?

- A) 6,8; B) 5; C) 6; D) 6,5.

Câu hỏi liên quan đến kỹ năng đọc thông tin từ bảng số liệu được cho dưới dạng bảng tần số và tính số trung bình cộng. Có 88,4% HS trả lời đúng câu hỏi này. Như vậy, hầu hết HS đọc hiểu được số liệu từ bảng, biết cách tính trung bình cộng và thực hiện các phép toán chính xác.

Câu 2: Một nhãn dán được in trên chai thuốc có lời cảnh báo sau: “Khi bôi thuốc lên da có 15% khả năng bị dị ứng nổi mẩn. Nếu bị dị ứng, hãy tham khảo ý kiến của bác sĩ”. Điều nào sau đây giải thích đúng nhất lời cảnh báo này?

- A) Không nên sử dụng loại thuốc này vì có nhiều khả năng bị dị ứng.
- B) Khi bôi lên da, chỉ áp dụng 15% liều lượng được đề nghị.
- C) Nếu bị dị ứng, chỉ có 15% diện tích bề mặt da bị nổi mẩn.
- D) Khoảng 15 trên 100 người sử dụng thuốc này bị dị ứng.

Câu hỏi yêu cầu HS giải thích ý nghĩa của lời cảnh báo ghi trên một lọ thuốc rằng có 15% nguy cơ bị phản ứng dị ứng da khi sử dụng sản phẩm này. Để hiểu đúng cảnh báo, HS cần nhận thức được rằng “15% nguy cơ dị ứng” có nghĩa là xác suất xảy ra phản ứng dị ứng khi dùng loại thuốc này là 15%; hay nói cách khác, trong 100 người sử dụng, khoảng 15 người sẽ bị dị ứng. Trong câu hỏi này, 79,5% HS đã hiểu đúng ý nghĩa của cảnh báo trên lọ thuốc và trả lời chính xác.

Hai câu hỏi có độ khó lớn nhất là câu 8 và câu 10.

Câu 8: Năm 2011, có thông tin cho rằng giá nhà trung bình ở Hà Nội đã tăng 20% trong một năm, trong khi trung vị của giá nhà lại giảm 2% trong cùng năm đó. Giải thích nào sau đây là hợp lý nhất cho thông tin này?

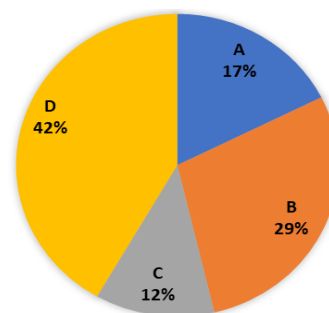
- A) Hầu hết các ngôi nhà ở Hà Nội giảm giá và một số ít ngôi nhà tăng giá mạnh.

B) Giá của hầu hết các ngôi nhà ở Hà Nội giảm và có nhiều ngôi nhà được bán trong năm 2011 hơn so với năm trước đó.

C) Thông tin chưa chính xác vì nếu giá nhà trung bình tăng thì trung vị giá nhà cũng phải tăng.

D) Giá của hầu hết các ngôi nhà ở Hà Nội tăng và có ít ngôi nhà được bán trong năm 2011 hơn so với năm trước đó.

Tỉ lệ HS lựa chọn các phương án như sau (xem biểu đồ 3). Để trả lời được câu hỏi này, HS cần biết được đặc điểm của số trung bình và trung vị. Giá trung bình được tính bằng cách lấy tổng giá trị của tất cả các ngôi nhà và chia cho số lượng ngôi nhà. Do đó, giá trung bình bị ảnh hưởng nhiều bởi những ngôi nhà có giá trị cao. Trung vị là giá trị nằm ở giữa khi sắp xếp giá trị của tất cả các ngôi nhà theo thứ tự từ thấp đến cao. Do đó, trung vị ít bị ảnh hưởng bởi những ngôi nhà có giá trị cao hay thấp. Từ đó có thể thấy một số ít ngôi nhà đã tăng giá rất mạnh, dẫn đến việc giá trung bình tăng cao, tuy nhiên nhiều ngôi nhà khác giảm giá dẫn đến việc trung vị giá nhà giảm. Chỉ có 17% HS trả lời đúng câu hỏi này.



Biểu đồ 3. Tỉ lệ lựa chọn các phương án của câu 8

Câu 10: 9 HS của một lớp học đã tiến hành cân một vật nhỏ trên cùng một cái cân và ghi lại khối lượng (tính bằng gam) của vật đó theo các giá trị như sau: 6,2; 6,0; 6,0; 15,3; 6,1; 6,2; 6,15; 6,2; 6,3. HS cần xác định chính xác khối lượng của vật này. Trong các phương pháp sau đây, các em nên sử dụng phương pháp nào?

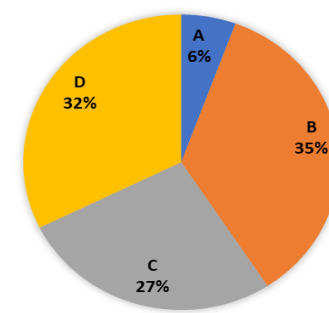
- A) Lấy giá trị 6,15 vì đó là số được lấy chính xác đến hàng phần trăm.
- B) Lấy trung bình của 9 giá trị.
- C) Loại trừ số 15,3, lấy trung bình 8 giá trị còn lại.
- D) Lấy giá trị xuất hiện nhiều nhất, đó là 6,2.

Tỉ lệ HS lựa chọn các phương án như trong biểu đồ 4. Đây là một tình huống HS có thể gặp trong thực tiễn. Để xác định chính xác khối lượng của vật, cần xem xét các phương pháp đề xuất, phân tích lí do hợp lí hoặc không hợp lí để đưa ra lựa chọn phù hợp.

Phương án A: Việc chọn một giá trị duy nhất mà không xem xét toàn bộ dữ liệu có thể dẫn đến sai sót nếu giá trị đó là một ngoại lệ, hoặc không đại diện cho toàn bộ dữ liệu.

Phương án B: Phương pháp này không hợp lí vì giá trị trung bình bị ảnh hưởng bởi giá trị ngoại lai là 15,3.

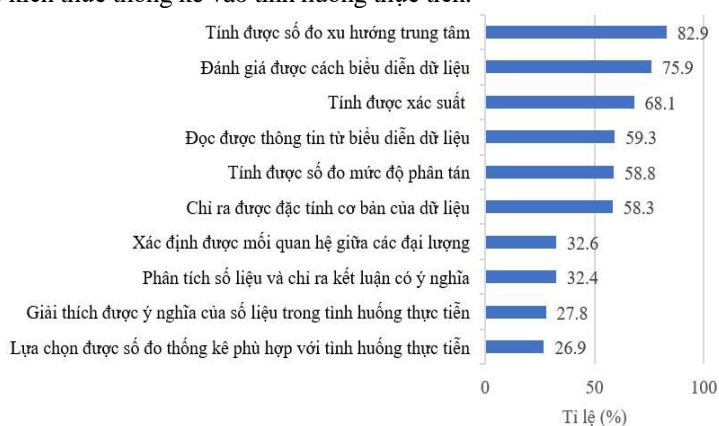
Phương án D: Phương pháp này không phù hợp vì không tận dụng được toàn bộ thông tin từ mẫu số liệu.



Biểu đồ 4. Tỉ lệ lựa chọn các phương án của câu 10

Như vậy, có 27% HS trả lời đúng câu 10 với phương án C. Đa số HS lựa chọn phương án B, trung bình cộng là một giá trị đại diện quen thuộc mà HS thường được yêu cầu tính toán trong các bài tập.

Tỉ lệ HS hoàn thành câu hỏi thuộc các dạng hoạt động cụ thể được tổng hợp trong hình 2. Qua đó, có thể thấy các hoạt động liên quan đến khả năng tính toán và kỹ năng cơ bản về thống kê có tỉ lệ hoàn thành cao: tính số đo xu hướng trung tâm (chiếm 82,9%); tính xác suất (chiếm 68,1%); đọc thông tin từ biểu diễn dữ liệu (chiếm 59,3%); tính được số đo mức độ phân tán (chiếm 58,8%). Các hoạt động cần khả năng phân tích, lí giải và vận dụng có tỉ lệ hoàn thành thấp hơn đáng kể: lựa chọn số đo thống kê phù hợp (chiếm 26,9%), giải thích ý nghĩa số liệu (chiếm 27,8%), phân tích số liệu và rút ra kết luận (chiếm 32,4%), xác định mối quan hệ giữa các đại lượng (chiếm 32,6%). Nhìn chung, kết quả cho thấy HS tương đối vững vàng về các kỹ năng cơ bản như tính toán, đọc biểu đồ nhưng còn hạn chế trong việc phân tích, vận dụng và lí giải các vấn đề liên quan đến số liệu thống kê. Đặc biệt, HS gặp khó khăn trong việc áp dụng các kiến thức thống kê vào tình huống thực tiễn.



Hình 2. Tỉ lệ HS hoàn thành các dạng câu hỏi trong bài kiểm tra

- Kiểm định sự khác biệt: Số liệu thu được sau khi phân tích bằng phần mềm QUEST được chuyển sang phần mềm SPSS để thực hiện phân tích tiếp theo. Kiểm định Independent-Samples T-Test được sử dụng để kiểm định sự khác biệt về điểm kiểm tra, khả năng SLTK theo giới tính. Kết quả được trình bày trong bảng 4.

Bảng 4. Kết quả kiểm định sự khác biệt trung bình theo giới tính

	Nhóm mẫu	Số lượng	Trung bình	Độ lệch chuẩn	Chênh lệch	t	sig
Điểm kiểm tra	Nữ	50	5,680	1,3952	-0,0269	-0,1	0,92
	Nam	58	5,707	1,3830			
Khả năng SLTK	Nữ	50	0,3588	0,72345	-0,00948	-0,068	0,946
	Nam	58	0,3683	0,72260			

Theo thông tin từ bảng 4, HS nam có trung bình điểm kiểm tra và khả năng SLTK cao hơn HS nữ, tuy nhiên sự chênh lệch là rất nhỏ. Kiểm định có sig > 0,05 nên chưa có cơ sở khẳng định sự khác biệt của hai nhóm đối tượng này.

3. Kết luận

Các câu hỏi nghiên cứu đặt ra đã được trả lời thông qua những nội dung phân tích trong mục 2.3.2. Cụ thể: Với câu hỏi nghiên cứu thứ nhất: Đề kiểm tra có độ tin cậy cao (Reliability of estimate đạt 0,95), sự phân bố về độ khó các câu hỏi trong đề thi bao trùm được khả năng của HS, kết quả đo được phản ánh đúng khả năng của HS tham gia làm bài. Như vậy, đề kiểm tra phù hợp để đánh giá khả năng SLTK của HS. Với câu hỏi nghiên cứu thứ hai: Khả năng SLTK của HS có phân bố khá đồng đều, 48% có khả năng trên mức trung bình. Trên thang đo logit, khả năng SLTK của HS đạt giá trị trong đoạn [-1,31; 2,05], giá trị trung bình là 0,36, độ lệch chuẩn là 0,72. HS thực hiện tương đối tốt các kỹ năng cơ bản như tính toán, đọc và hiểu biểu đồ trình bày số liệu. Tuy nhiên, HS còn gặp hạn chế trong việc vận dụng kiến thức thống kê để phân tích, lí giải và rút ra kết luận có ý nghĩa từ các số liệu đã cho. HS thể hiện khó khăn khi phải liên hệ, áp dụng các khái niệm và công cụ thống kê vào các tình huống thực tiễn. Với câu hỏi nghiên cứu thứ ba: Không có sự khác biệt về khả năng SLTK giữa nhóm HS nam và HS nữ. Điều này cũng đồng nhất kết quả nghiên cứu của Liu và Garfield (2002). Trong khi một số nghiên cứu đã chỉ ra HS nam có thái độ đối với TK&XS tích cực hơn, thể hiện sự hứng thú và sự tự tin hơn so với HS nữ. Ngược lại, HS nữ có xu hướng cho rằng TK&XS là một chủ đề khó khăn hơn so với HS nam (Chiesi & Primi, 2015; Chowdhury, 2018). Tuy nhiên,

trong nghiên cứu này của chúng tôi không tìm thấy sự khác biệt về khả năng SLTK hay sự khác biệt về thành tích học tập TK&XS giữa hai nhóm đối tượng này.

Kết quả của nghiên cứu của bài báo phản ánh thực trạng về khả năng SLTK của HS THPT. Điều này gợi ra những điều chỉnh và trọng tâm cần lưu ý cho GV trong dạy học TK&XS: Nhấn mạnh sự cần thiết phải tăng cường rèn luyện khả năng vận dụng, phân tích và lí giải số liệu thống kê cho HS. Mặc dù HS có thể thực hiện tương đối tốt các kĩ năng cơ bản, nhưng cần giúp các em biết cách ứng dụng những kiến thức đó vào thực tiễn; cần thiết kể các dạng bài tập tình huống, bài tập ứng dụng để HS được luyện tập xử lí, phân tích số liệu trong các tình huống gần gũi với cuộc sống; khuyến khích HS đưa ra các giải pháp, kết luận dựa trên phân tích số liệu thay vì chỉ thực hiện các phép tính đơn giản hay trình bày số liệu.

Tài liệu tham khảo

- Adams, R. J., & Khoo, S.-T. (1996). *Quest: The interactive test analysis system*. Melbourne ACER.
- Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation.
- Chiesi, F., & Primi, C. (2015). *Gender differences in attitudes toward statistics: Is there a case for a confidence gap?* CERME 9-Ninth congress of the European society for research in mathematics education, Charles University in Prague, Faculty of Education; ERME, Feb 2015, Prague, Czech Republic, pp. 622-628.
- Chowdhury, S. (2018). Prior knowledge, sex and students' attitude towards statistics: a study on postgraduate education students. *American Journal of Educational Research*, 6(3), 270-276.
- Delmas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28-58.
- Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22-38.
- Garfield, J. B., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.
- Garfield, J., & Ben-Zvi, D. (2008). *Preparing school teachers to develop students' statistical reasoning*. Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study: The 18th ICMI Study, 299-310.
- Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking Learning*, 2(1-2), 99-125.
- Georg, R. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Institute of Education Research.
- Hoàng Lê Minh (2023). *Xây dựng môi trường học tập nội dung thống kê ở trường trung học phổ thông theo hướng rèn luyện kĩ năng suy luận thống kê cho học sinh*. Luận án tiến sĩ Khoa học Giáo dục, Trường Đại học Vinh.
- Hoàng Nam Hải (2013). *Phát triển năng lực suy luận thống kê cho sinh viên cao đẳng chuyên nghiệp*. Luận án tiến sĩ Giáo dục học, Trường Đại học Vinh.
- Lâm Quang Thiệp (2010). *Đo lường trong giáo dục, lí thuyết và ứng dụng*. NXB Đại học Quốc gia Hà Nội.
- Liu, H.-J. C., & Garfield, J. B. (2002). Sex differences in statistical reasoning. *Bulletin of Educational Psychology*, 34(1), 123-137.
- Morales, R. A. (2009). Evaluation of mathematics achievement test: A comparison between CTT and IRT. *The International Journal of Educational Psychological Assessment*, 1(1), 19-26.
- Quách Thị Sen (2020). *Phát triển tư duy thống kê cho sinh viên đại học ngành dược trong dạy học Toán - thống kê y dược*. Luận án tiến sĩ Khoa học Giáo dục, Trường Đại học Sư phạm Hà Nội.
- Sabbag, A. G., & Zieffler, A. (2015). Assessing learning outcomes: An analysis of the GOALS-2 instrument. *Statistics Education Research Journal*, 14(2), 93-116.
- Trần Đức Chiển (2007). *Rèn luyện tư duy thống kê cho học sinh trong dạy học thống kê - xác suất ở môn Toán trung học phổ thông*. Luận án tiến sĩ Giáo dục học, Viện Khoa học Giáo dục Việt Nam.