

ỨNG DỤNG LÝ THUYẾT TRẮC NGHIỆM CỔ ĐIỂN TRONG PHÂN TÍCH CÂU HỎI TRẮC NGHIỆM KHÁCH QUAN

• Nguyễn Trung Hiếu^(*), Nguyễn Bích Như^(*)

Tóm tắt

Bài viết báo cáo kết quả nghiên cứu và vận dụng Lý thuyết trắc nghiệm cổ điển trong phân tích câu hỏi trắc nghiệm khách quan. Phương pháp thống kê toán học được sử dụng để phân tích một bộ đề trắc nghiệm khách quan 30 câu hỏi với 85 bài làm của sinh viên. Kết quả cho thấy lý thuyết này có thể được vận dụng để phân tích các thông số của câu hỏi trắc nghiệm như độ khó, độ phân biệt, chất lượng của các phương án nhiễu... Các thông số này đóng vai trò quan trọng trong việc chuẩn hóa và nâng cao dần chất lượng của các câu hỏi thi. Việc sử dụng các câu hỏi đã được chuẩn hóa giúp cho hoạt động kiểm tra, đánh giá kết quả học tập trong nhà trường được chính xác, khách quan và công bằng hơn.

Từ khóa: Lý thuyết trắc nghiệm cổ điển, câu hỏi trắc nghiệm khách quan, đánh giá kết quả học tập.

1. Đặt vấn đề

Hướng tới yêu cầu đánh giá công bằng, khách quan kết quả học tập của người học, việc đa dạng hóa các hình thức kiểm tra đang được triển khai rộng rãi. Trong đó, phải kể đến là hình thức trắc nghiệm khách quan. Đây là phương pháp kiểm tra, đánh giá kết quả học tập của người học bằng hệ thống các câu hỏi trắc nghiệm (CHTN). Tuy nhiên, đa số đề thi trắc nghiệm được sử dụng ở các trường hiện nay đều chưa được thử nghiệm và tu chỉnh. Chất lượng của những đề thi như thế vẫn còn là một ẩn số. Chính vì vậy, việc thiết kế và xây dựng nên những bộ đề trắc nghiệm dạng tiêu chuẩn hóa đang trở thành một yêu cầu bức thiết. Các CHTN tiêu chuẩn hóa được soạn thảo, thử nghiệm và tu chỉnh một cách chi tiết, tỉ mỉ. Mỗi một câu hỏi đều phải được phân tích để xem xét các thuộc tính về độ khó, độ phân biệt cũng như các tham số cần thiết khác. Cơ sở khoa học để thực hiện những phân tích trên phải kể đến đầu tiên là Lý thuyết trắc nghiệm cổ điển (Classical Test Theory). Mặc dù xu thế phát triển hiện nay của khoa học đo lường, đánh giá trên thế giới là thiên về các lý thuyết đo lường hiện đại (có thể kể đến là lý thuyết Ứng đáp câu hỏi - Item Response Theory); tuy nhiên, ở Việt Nam, có một thực tế là việc tiếp cận các học thuyết đo lường đánh giá nói chung vẫn còn chậm hơn nhiều so với thế giới. Theo giáo sư Lâm Quang Thiệp: Trong nhà trường, kể cả các trường đại học sư phạm, cũng chưa có các chương trình đào tạo về lĩnh vực này một cách bài bản... Tình hình nói trên đòi hỏi

chúng ta phải có nhiều tài liệu giới thiệu để một mặt nâng cao hiểu biết về phương pháp này cho đội ngũ nhà giáo cũng như công chúng, mặt khác, góp phần thúc đẩy sự hình thành lĩnh vực khoa học về đo lường trong giáo dục ở nước ta” [4].

Chính vì vậy, để đưa khoa học này từng bước phát triển trong nhà trường nói riêng và xã hội nói chung, Lý thuyết trắc nghiệm cổ điển chính là sự lựa chọn phù hợp đầu tiên để triển khai; bởi đây là lý thuyết nền tảng của khoa học đo lường, đánh giá. Đó cũng là cơ sở vững chắc để giáo viên tiếp cận tốt hơn với lý thuyết đo lường hiện đại, bắt kịp xu hướng chung của thế giới. Bài viết báo cáo kết quả phân tích các CHTN của một đề kiểm tra dựa trên nền tảng của Lý thuyết trắc nghiệm cổ điển nhằm góp phần khẳng định tính khả thi và hiệu quả của việc vận dụng lý thuyết này vào trong thực tế kiểm tra, đánh giá ở các trường hiện nay.

2. Đôi nét về Lý thuyết trắc nghiệm cổ điển

Lý thuyết trắc nghiệm cổ điển (gọi tắt là CTT) còn được biết đến với tên là Lý thuyết về điểm số thực (true score theory), chỉ việc phân tích kết quả các bài kiểm tra dựa trên điểm số [5]. Lý thuyết CTT cho rằng năng lực của mỗi thí sinh được xác định bởi một điểm thực T. Tuy nhiên, trong thực tế, điểm thực không bao giờ có thể thu được một cách trực tiếp mà chỉ có thể là một điểm quan sát X nào đó. Vì vậy, phương trình cơ bản của Lý thuyết CTT có dạng [1]:

$$X = T + E$$

Trong đó:

X là điểm quan sát (điểm làm bài của thí sinh)

T là điểm thực của thí sinh

E là sai số

^(*) Trường Cao đẳng Sư phạm Sóc Trăng.

Lý thuyết trắc nghiệm cổ điển quan tâm đến các tham số sau đây [5]:

Độ khó của CHTN (giá trị p): được xác định bằng tỷ số phần trăm thí sinh làm đúng câu hỏi trên tổng số thí sinh tham gia làm câu hỏi đó.

$$p_i = \frac{n}{N} \quad (1)$$

Trong đó:

p_i là độ khó của câu hỏi thứ i

n là số thí sinh làm đúng câu hỏi thứ i

N là tổng số thí sinh tham gia làm câu hỏi thứ i

Theo Osterlind (1989), độ khó của câu hỏi nên nằm trong khoảng từ 0,4 đến 0,8 [3]. Câu hỏi có độ khó lớn hơn 0,8 là quá dễ; có độ khó nhỏ hơn 0,4 là quá khó.

Để tính độ khó trung bình (p_{tb}) của một câu hỏi cần phải căn cứ số phương án trả lời của câu hỏi đó. Về lý thuyết, độ khó trung bình của một câu hỏi có k phương án chọn là: [4]

$$p_{tb} = \frac{100\% + \frac{1}{k}}{2} \quad (2)$$

Trong đó:

p_{tb} là độ khó trung bình của câu hỏi

k là số phương án trả lời có trong câu hỏi

Ngoài ra, giá trị p còn giúp chỉ ra một số lỗi của câu hỏi như: nhằm đáp án, lỗi do dùng từ, ngữ pháp làm thí sinh không hiểu hoặc hiểu nhầm... Khi đó, giá trị p sẽ có sự bất thường (có thể là đạt quá cao ở phương án nhiều và quá thấp ở phương án đúng...).

Độ phân biệt của CHTN (giá trị D): là khả năng của CHTN thực hiện được sự phân biệt nhóm thí sinh thành các nhóm năng lực khác nhau (như giỏi, trung bình, kém...). Theo Lâm Quang Thiệp (2008) [4], để tính độ phân biệt, người ta dựa vào

tổng điểm thô của từng thí sinh để tách nhóm thí sinh thành 2 nhóm: một nhóm giỏi (gồm 27% thí sinh đạt điểm cao từ trên xuống) và một nhóm kém (gồm 27% thí sinh đạt điểm thấp từ dưới lên). Khi đó, biểu thức tính độ phân biệt của một CHTN được viết như sau:

$$D_i = \frac{C - T}{S} \quad (3)$$

Trong đó:

D_i là độ phân biệt của câu hỏi thứ i

C là số thí sinh làm đúng câu hỏi thứ i thuộc nhóm giỏi

T là số thí sinh làm đúng câu hỏi thứ i thuộc nhóm kém

S là số lượng thí sinh của một trong hai nhóm (27% tổng số)

Ebel (1965) cho rằng các câu hỏi của bài thi nên có độ phân biệt lớn hơn hoặc bằng 0,3 [2].

Bên cạnh độ khó và độ phân biệt của CHTN, lý thuyết CTT còn quan tâm đến 2 đại lượng khác là độ tin cậy và độ giá trị của đề trắc nghiệm. Tuy nhiên, bài viết này chỉ bàn sâu về các đại lượng gắn liền với CHTN là độ khó và độ phân biệt của CHTN.

3. Ứng dụng Lý thuyết trắc nghiệm cổ điển trong phân tích CHTN

Dữ liệu được dùng để phân tích là 30 CHTN (4 phương án lựa chọn) trong bài kiểm tra môn Đại số tuyến tính do 85 sinh viên ở 3 lớp Toán - Lý, Tin - Lý, Tin học Khóa 16 của Trường Cao đẳng Sư phạm Sóc Trăng thực hiện. Dữ liệu được nhập vào chương trình bảng tính Excell và kiểm tra hai lần để đảm bảo độ chính xác.

3.1. Kết quả phân tích độ khó

Độ khó (p) của từng CHTN trong bài kiểm tra này được tính toán theo công thức (1). Kết quả chi tiết được trình bày trong Bảng 1.

Bảng 1. Tổng hợp độ khó của 30 câu hỏi trắc nghiệm

CHTN	Câu 6	Câu 26	Câu 27	Câu 22	Câu 1	Câu 19	Câu 14	Câu 25	Câu 4	Câu 9
p	0,91	0,91	0,89	0,88	0,87	0,87	0,86	0,86	0,85	0,78
CHTN	Câu 13	Câu 28	Câu 5	Câu 8	Câu 11	Câu 30	Câu 7	Câu 3	Câu 12	Câu 18
p	0,78	0,76	0,72	0,72	0,72	0,71	0,69	0,66	0,65	0,65
CHTN	Câu 29	Câu 20	Câu 15	Câu 24	Câu 21	Câu 10	Câu 23	Câu 2	Câu 17	Câu 16
p	0,62	0,60	0,53	0,47	0,42	0,39	0,36	0,35	0,34	0,33

Bảng 1 cho thấy độ khó của các CHTN (được sắp xếp theo thứ tự từ lớn đến bé) trong bài kiểm tra này nằm trong khoảng từ 0,33 đến 0,91, trong đó:

$p < 0,4$ có 5 câu gồm: câu 16, 17, 2, 23 và câu 10, chiếm 17%. Đây là các câu hỏi khó so với năng lực của sinh viên. Khó nhất là câu 16, chỉ có 33% sinh viên trả lời đúng. Các câu này vượt ra ngoài độ khó giới hạn theo khuyến nghị của Osterlind (1989) [3] nên nếu có thể, cần được điều chỉnh lại cho phù hợp. Tuy nhiên, cần lưu ý là bất kỳ đề kiểm tra nào cũng phải đảm bảo số lượng các câu hỏi khó theo quy định để đánh giá được sinh viên ở mức năng lực cao.

$0,4 \leq p \leq 0,8$ có 16 câu, chiếm phần lớn (53%) các câu hỏi trong đề kiểm tra này. Đó là các câu 9, 13, 28, 5, 8, 11, 30, 7, 3, 12, 18, 29, 20, 15, 24, 21. Theo Osterlind (1989) [3], đây là những câu hỏi có độ khó phù hợp.

$p > 0,8$ có 9 câu chiếm 30% các câu hỏi trong đề kiểm tra. Đó là các câu 6, 26, 27, 22, 1, 19, 14, 25, 4. Đây là các câu hỏi dễ so với năng lực của sinh viên. Đặc biệt là câu 6 và 26 có trên 90% sinh viên trả lời đúng. Đề thi cần có câu hỏi dễ để đánh giá năng lực của những sinh viên yếu nhưng không nên có quá nhiều như đề kiểm tra này. Vì vậy, các câu hỏi này cần được thiết kế lại cho phù hợp.

Bảng 2. Tổng hợp các giá trị chênh lệch giữa nhóm trên và nhóm dưới ở từng phương án

	Câu 1	Câu 2	Câu 3	Câu 4	Câu 5	Câu 6	Câu 7	Câu 8	Câu 9	Câu 10
Đáp án	B	D	B	D	D	B	B	A	B	B
Chênh lệch tỷ lệ ở phương án A	0,00	-0,35	-0,26	-0,09	-0,04	0,00	-0,17	0,43	-0,13	-0,13
Chênh lệch tỷ lệ ở phương án B	0,13	-0,04	0,70	0,00	-0,17	0,13	0,57	-0,17	0,30	0,57
Chênh lệch tỷ lệ ở phương án C	-0,13	-0,09	-0,22	-0,09	-0,22	-0,13	-0,22	-0,17	-0,13	-0,30
Chênh lệch tỷ lệ ở phương án D	0,00	0,48	-0,22	0,17	0,43	0,00	-0,17	-0,09	-0,04	-0,09
	Câu 11	Câu 12	Câu 13	Câu 14	Câu 15	Câu 16	Câu 17	Câu 18	Câu 19	Câu 20
Đáp án	A	A	D	B	A	B	B	B	A	C
Chênh lệch tỷ lệ ở phương án A	0,65	0,65	-0,22	-0,09	0,74	0,00	-0,26	-0,17	0,43	-0,04
Chênh lệch tỷ lệ ở phương án B	-0,30	0,00	-0,13	0,17	-0,30	0,35	0,43	0,48	-0,22	0,04

Thêm vào đó, độ khó trung bình (p_{tb}) của các CHTN (4 phương án) trong đề kiểm tra này cũng được tính toán theo công thức (2):

$$p_{tb} = \frac{100\% + \frac{1}{k}}{4} = \frac{100\% + \frac{1}{4}}{4} = 0,625 \approx 0,63$$

Đối chiếu với các kết quả tính toán về độ khó ở Bảng 1, ta thấy, đề kiểm tra này có đến 2/3 số câu (66,7%) có độ khó lớn hơn độ khó trung bình và chỉ có 1/3 số câu (33,3%) có độ khó nhỏ hơn độ khó trung bình. Như vậy, nhìn chung, có thể nói đề kiểm tra này có nhiều câu dễ so với năng lực của sinh viên.

3.2. Kết quả phân tích độ phân biệt

Việc tính toán độ phân biệt của từng CHTN được tiến hành theo trình tự các bước như Lâm Quang Thiệp (2008) [4] đã đề xuất. Đầu tiên, ta cần dựa trên tổng điểm bài làm của sinh viên để tách thành 2 nhóm. Số lượng mỗi nhóm là 23 sinh viên (27% trong tổng số 85 sinh viên). Sau khi sắp xếp lại dữ liệu bài làm theo 2 nhóm, ta bắt đầu tính toán các giá trị chênh lệch giữa 2 nhóm ở từng phương án theo công thức (3). Khi đó, độ phân biệt (D) chính là kết quả sự chênh lệch giữa hai nhóm ngay tại phương án đúng. Kết quả chi tiết được trình bày trong Bảng 2.

Chênh lệch tỷ lệ ở phương án C	-0,26	-0,13	-0,22	-0,04	-0,30	-0,13	-0,04	-0,13	-0,13	0,09
Chênh lệch tỷ lệ ở phương án D	-0,09	-0,52	0,57	-0,04	-0,13	-0,22	-0,13	-0,17	-0,09	-0,09
	Câu 21	Câu 22	Câu 23	Câu 24	Câu 25	Câu 26	Câu 27	Câu 28	Câu 29	Câu 30
Đáp án	C	B	A	C	C	A	B	B	B	C
Chênh lệch tỷ lệ ở phương án A	-0,39	-0,13	0,65	-0,09	-0,26	0,09	-0,04	-0,17	-0,17	0,00
Chênh lệch tỷ lệ ở phương án B	0,00	0,22	-0,35	-0,30	-0,13	-0,04	0,30	0,48	0,39	-0,17
Chênh lệch tỷ lệ ở phương án C	0,48	0,00	-0,17	0,70	0,52	0,00	-0,17	-0,04	-0,13	0,35
Chênh lệch tỷ lệ ở phương án D	-0,09	-0,09	-0,13	-0,30	-0,13	-0,04	-0,09	-0,26	-0,09	-0,17

Bảng 3. Tổng hợp độ phân biệt của câu hỏi

	Câu 1	Câu 2	Câu 3	Câu 4	Câu 5	Câu 6	Câu 7	Câu 8	Câu 9	Câu 10
Đáp án	B	D	B	D	D	B	B	A	B	B
Độ phân biệt	0,13	0,48	0,70	0,17	0,43	0,13	0,57	0,43	0,30	0,57
	Câu 11	Câu 12	Câu 13	Câu 14	Câu 15	Câu 16	Câu 17	Câu 18	Câu 19	Câu 20
Đáp án	A	A	D	B	A	B	B	B	A	C
Độ phân biệt	0,65	0,65	0,57	0,17	0,74	0,35	0,43	0,48	0,43	0,09
	Câu 21	Câu 22	Câu 23	Câu 24	Câu 25	Câu 26	Câu 27	Câu 28	Câu 29	Câu 30
Đáp án	C	B	A	C	C	A	B	B	B	C
Độ phân biệt	0,48	0,22	0,65	0,70	0,52	0,09	0,30	0,48	0,39	0,35

Theo Bảng 3, ta thấy hầu hết các câu trong đề kiểm tra này đều đạt độ phân biệt tốt ($\geq 0,3$). Tuy nhiên, vẫn còn 7 câu (chiếm 23%) có độ phân biệt thấp. Đó là câu 1, câu 4, câu 6, câu 14, câu 20, câu 22, câu 26. Những câu này cần phải được điều chỉnh để đạt độ phân biệt tốt hơn.

3.3. Các kết quả phân tích khác

Khả năng nhầm đáp án: Như đã nói, việc

nhầm lẫn có thể dễ dàng bị phát hiện khi xem xét giá trị p: có sự khác biệt lớn giữa phương án đúng (p thường nhỏ hơn rất nhiều) so với các phương án còn lại. Bảng 4 cho thấy, các câu hỏi trong đề kiểm tra này không có dấu hiệu của việc nhầm đáp án. Các phương án đúng đều có tỉ lệ sinh viên chọn cao hơn các phương án nhiễu.

Bảng 4. Tỷ lệ lựa chọn các phương án của từng câu hỏi trắc nghiệm

Tỷ lệ chọn	Câu 1	Câu 2	Câu 3	Câu 4	Câu 5	Câu 6	Câu 7	Câu 8	Câu 9	Câu 10
A	0,05	0,24	0,13	0,06	0,08	0,01	0,11	0,72	0,08	0,18
B	0,87	0,18	0,66	0,05	0,09	0,91	0,69	0,08	0,78	0,39

C	0,06	0,24	0,09	0,05	0,11	0,07	0,08	0,09	0,08	0,24
D	0,02	0,35	0,12	0,85	0,72	0,01	0,12	0,11	0,06	0,19
Độ khó p	0,87	0,35	0,66	0,85	0,72	0,91	0,69	0,72	0,78	0,39
Tỷ lệ chọn	Câu 11	Câu 12	Câu 13	Câu 14	Câu 15	Câu 16	Câu 17	Câu 18	Câu 19	Câu 20
A	0,72	0,65	0,07	0,05	0,53	0,21	0,24	0,13	0,87	0,16
B	0,08	0,04	0,06	0,86	0,14	0,33	0,34	0,65	0,07	0,14
C	0,12	0,05	0,09	0,05	0,18	0,25	0,22	0,09	0,04	0,60
D	0,08	0,27	0,78	0,05	0,15	0,21	0,20	0,13	0,02	0,09
Độ khó p	0,72	0,65	0,78	0,86	0,53	0,33	0,34	0,65	0,87	0,60
Tỷ lệ chọn	Câu 21	Câu 22	Câu 23	Câu 24	Câu 25	Câu 26	Câu 27	Câu 28	Câu 29	Câu 30
A	0,21	0,04	0,36	0,14	0,07	0,91	0,04	0,07	0,11	0,11
B	0,18	0,88	0,24	0,11	0,04	0,05	0,89	0,76	0,62	0,07
C	0,42	0,06	0,16	0,47	0,86	0,02	0,05	0,06	0,11	0,71
D	0,19	0,02	0,24	0,27	0,04	0,02	0,02	0,11	0,16	0,12
Độ khó p	0,42	0,88	0,36	0,47	0,86	0,91	0,89	0,76	0,62	0,71

Bảng 5. Đề xuất việc điều chỉnh một số phương án nhiều

Tỷ lệ chọn	Câu 6	Câu 12	Câu 20	Câu 23	Câu 24
A	0,01	0,65	0,16	0,36	0,14
B	0,91	0,04	0,14	0,24	0,11
C	0,07	0,05	0,6	0,16	0,47
D	0,01	0,27	0,09	0,24	0,27
Đáp án	B	A	C	A	C
Phương án nhiều cần điều chỉnh	A, D	D	D	C	A, B

Chất lượng của các phương án nhiều: Chất lượng phương án nhiều được xem là lý tưởng khi tỉ lệ sinh viên lựa chọn chúng tương đối đồng đều. Các phương án bị bỏ qua hoặc chỉ có một số ít lựa chọn chứng tỏ đó là những phương án sai quá lộ liễu. Như vậy, nếu các phương án nhiều có tỉ lệ lựa chọn quá chênh lệch thì cần xem xét và thiết kế lại chúng. Bảng 4 cho thấy: Trong đề kiểm tra này có nhiều câu đạt yêu cầu về phương án nhiều. Tỉ lệ sinh viên lựa chọn chúng khá đồng đều. Điển hình là câu 1 (0,05: 0,06: 0,02), câu 3 (0,13 : 0,09

: 0,12), câu 15 (0,14 : 0,18 : 0,15), Ngoài ra còn có câu 7, 8, 9, 11, 13... Tuy nhiên, đề kiểm tra này vẫn còn một số câu cần được điều chỉnh ở phương án nhiều chẳng hạn như câu 6, 12, 20, 23, 24 (Bảng 5).

4. Kết luận

Tóm lại, trên cơ sở vững chắc của Lý thuyết trắc nghiệm cổ điển, bài viết đã tiến hành các bước thực nghiệm để phân tích thuộc tính của các CHTN có trong một đề kiểm tra. Kết quả đã đánh giá được các thông số về độ khó, độ phân biệt cũng như các vấn đề liên quan đến CHTN như chất lượng các phương án nhiều, khả năng nhầm đáp án. Bài viết một lần nữa khẳng định tính khả thi và hiệu quả của việc vận dụng lý thuyết trắc nghiệm cổ điển vào trong thực tế kiểm tra, đánh giá ở các trường. Thông qua kết quả có được từ việc phân tích, giáo viên, nhà quản lý có thể đưa ra những kiến nghị chính xác, tin cậy nhằm góp phần nâng cao chất lượng câu hỏi thi. Nhìn xa hơn, việc sử dụng những đề thi với các câu hỏi đã được tu chỉnh, chuẩn hóa sẽ đánh giá được chính xác năng lực thực sự của người học; từ đó góp phần nâng cao chất lượng giáo dục của nhà trường./.

Tài liệu tham khảo

- [1]. Robert L. Brennan (2011), “Generalizability Theory and Classical Test Theory”, *Applied Measurement in Education*, 24: 1-21.
- [2]. Ebel, R. L. (1965), *Measuring Educational Achievement*, Englewood Cliffs: Prentice - Hall.
- [3]. Osterlind, S. J. (1989), *Constructing test items*, Boston: Kluwer Academic.
- [4]. Lâm Quang Thiệp (2008), *Trắc nghiệm và ứng dụng*, NXB Khoa học và Kỹ thuật, Hà Nội.
- [5]. Margaret Wu, Hak Ping Tam, Tsung-Hau Jen (2016), *Educational Measurement for Applied Researchers: Theory into Practice*, Springer Nature Singapore Pte Ltd.

**USING CLASSICAL TEST THEORY TO ANALYZE
MULTIPLE - CHOICE QUESTIONS****Summary**

The article reports the result of using Classical Test Theory to analyze multiple - choice questions. The statistical method was used to examine an already-used test of 30 items with 85 student responses. The findings indicate that Classical Test Theory can be used to analyze test item parameters such as item difficulty, item discrimination, and distractor quality. These test item parameters play an important role in standardizing and enhancing the quality of multiple - choice questions. Standardized tests ensure the accuracy, objectiveness and fairness in assessing student learning outcomes.

Keywords: Classical Test Theory, multiple-choice questions, assessing student learning outcomes.
Ngày nhận bài: 21/11/2017; Ngày nhận lại: 28/3/2018; Ngày duyệt đăng: 03/12/2018.