

# MỘT SỐ VẤN ĐỀ VỀ THỬ NGHIỆM CÂU HỎI KIỂM TRA

• ThS. **DẶNG XUÂN CƯỜNG**

*Dự án Phát triển Giáo dục Trung học cơ sở II*

**Đ**ánh giá chất lượng giáo dục nói chung và đánh giá chất lượng học tập của học sinh (HS) nói riêng là một trong những nội dung quan trọng trong quá trình dạy học. Việc đánh giá đúng chất lượng học tập của HS sẽ giúp cho HS thấy được kết quả học tập của cá nhân, từ đó có các biện pháp phấn đấu, điều chỉnh quá trình học tập của bản thân, đồng thời cung cấp các thông tin hữu ích giúp giáo viên và các cấp quản lí có các điều chỉnh phù hợp nhằm đạt được mục tiêu giáo dục đề ra.

Đề kiểm tra là một trong những công cụ chủ yếu nhằm đánh giá kết quả học tập của HS, do đó việc thiết kế một đề kiểm tra có chất lượng, đảm bảo độ tin cậy là một trong những yêu cầu cấp thiết, đặc biệt là đối với những đợt kiểm tra quan trọng như kiểm tra hết chương, kiểm tra học kì hoặc các bài khảo sát kết quả học tập,... Tuy nhiên, một đề kiểm tra tốt, trước hết các câu hỏi trong đề kiểm tra đó phải có chất lượng.

Qua một cuộc khảo sát tại một số cơ sở giáo dục, đứng trước các đợt kiểm tra quan trọng, để thiết kế một đề kiểm tra cho HS, các cơ sở giáo dục thường tiến hành việc ra đề tập trung, trong đó dựa vào kiến thức, vốn kinh nghiệm giảng dạy của mình, giáo viên sưu tầm hoặc biên soạn mới các câu hỏi để thiết kế đề kiểm tra đó.

Với cách làm như trên, rõ ràng có hạn chế như: (i) Chất lượng các câu hỏi trong đề kiểm tra không được đảm bảo, các câu hỏi có thể có sai sót hoặc có lỗi về mặt kĩ thuật; (ii) Các đề kiểm tra không bám sát chuẩn kiến thức, kĩ năng đã được quy định; (iii) Giáo viên, sau khi kiểm tra lại kết quả làm bài sẽ không đánh giá đúng được năng lực thực sự của mỗi HS, bên cạnh đó sẽ không thu được các thông tin cần thiết để điều chỉnh quá trình dạy học của mình.

Do đó, vấn đề được đặt ra là, làm thế nào để các cơ sở giáo dục có thể nâng cao chất lượng

của các câu hỏi để xây dựng các bài kiểm tra phù hợp. Có nhiều cách thực hiện công việc này, tuy nhiên trong bài báo này, chúng tôi đưa ra một phương pháp, đó là tiến hành thử nghiệm các câu hỏi kiểm tra, đồng thời phân tích kết quả thu được nhằm điều chỉnh các câu hỏi để giảm thiểu tối đa các sai sót có thể xảy ra đối với câu hỏi đã biên soạn.

*Việc thử nghiệm các câu hỏi kiểm tra có một ý nghĩa quan trọng, nhằm các mục đích sau:*

- Kiểm tra lại tính chính xác về mặt kiến thức của các câu hỏi;
- Phát hiện ra các lỗi về mặt kĩ thuật viết câu hỏi;
- Xem xét mức độ phù hợp của các câu hỏi so với trình độ học sinh thông qua các chỉ số về độ khó, độ phân biệt của từng câu hỏi và của đề kiểm tra để điều chỉnh cho hợp lí;
- Kiểm nghiệm lại các thông tin dự đoán của các tác giả về các câu hỏi mà mình biên soạn;
- Xác định các tham số đặc trưng của các câu hỏi phục vụ việc so bằng (equating) các bài kiểm tra và phục vụ cho việc xây dựng ngân hàng câu hỏi.

Để làm rõ hơn vai trò của việc thử nghiệm, chúng tôi đã tiến hành phân tích một bài kiểm tra trắc nghiệm khách quan đã tiến hành vào cuối học kỳ I vừa qua. Đề kiểm tra bao gồm 30 câu trắc nghiệm khách quan (loại có 4 phương án lựa chọn) được ra cho 121 em HS lớp 6 trong thời gian 45 phút. Sau khi tiến hành kiểm tra, chúng tôi lập file dữ liệu các kết quả làm bài của HS đối với từng câu hỏi và đối với cả bài kiểm tra.

Trong phần tiếp theo, chúng tôi trình bày một số phương pháp để nâng cao chất lượng câu hỏi sau khi chúng đã được thử nghiệm.

## 1. Căn cứ vào độ khó của câu hỏi

Trong lí thuyết trắc nghiệm cổ điển, độ khó

của một câu hỏi là tỉ lệ giữa số thí sinh làm đúng câu hỏi trên tổng số thí sinh tham gia trả lời câu hỏi đó. Trong "Trắc nghiệm và ứng dụng" (Lâm Quang Thiệp, NXB Khoa học và Kỹ thuật, 2008) đã đưa ra một thang các mức độ về độ khó có thể chấp nhận được. Cụ thể, kí hiệu  $p$  là độ khó của câu hỏi, khi đó:

Nếu  $p > 0,75$  thì câu hỏi quá dễ;

Nếu  $p < 0,25$  thì câu hỏi quá khó;

Nếu  $0,25 \leq p \leq 0,75$  thì câu hỏi có độ khó chấp nhận được.

Việc tính toán độ khó của từng câu hỏi theo công thức trên là đơn giản và giáo viên có thể tiến hành một cách dễ dàng.

Có một điểm cần lưu ý khi câu hỏi có độ khó bằng 1, nghĩa là 100% học sinh tham gia đều trả lời đúng và câu hỏi có độ khó bằng 0, nghĩa là 100% học sinh tham gia đều trả lời sai. Đối với các câu hỏi này, tùy vào mục đích của kiểm tra, cần chỉnh sửa hoặc loại bỏ ra khỏi đề kiểm tra và thay thế chúng bằng các câu hỏi khác. Các câu hỏi này có thể chỉ phù hợp trong các bài kiểm tra theo tiêu chí khi cần kiểm tra trình độ đạt chuẩn đến mức độ nào, còn đối với các bài kiểm tra theo chuẩn thì không nên để các câu hỏi này vào vì như thế có nghĩa là các câu hỏi này không cung cấp thông tin gì để phân loại các thí sinh.

## 2. Căn cứ vào độ phân biệt của câu hỏi

Theo Dương Thiệu Tống ("Trắc nghiệm và đo lường thành quả học tập (phương pháp thực hành), NXB Khoa học Xã hội, 2005), độ phân biệt của câu hỏi là khả năng thực hiện sự phân biệt các thí sinh trong nhóm thí sinh tham gia trả lời câu hỏi đó.

Để tính độ phân biệt của câu hỏi, người ta thường phân loại căn cứ vào điểm thô của các thí sinh. Cụ thể, người ta phân loại nhóm giỏi gồm 27% tổng số thí sinh có điểm cao từ trên xuống và nhóm kém bao gồm 27% tổng số thí sinh có điểm thấp từ dưới lên. Gọi  $C$  là tổng số thí sinh trả lời đúng câu hỏi thuộc nhóm giỏi,  $T$  là tổng số thí sinh trả lời đúng câu hỏi thuộc nhóm kém và  $S$  là số lượng thí sinh của một trong hai nhóm trên. Khi đó, nếu  $D$  là độ phân

biệt của câu hỏi thì:

$$D = \frac{C - T}{S}$$

Theo công thức này, chúng tôi tiến hành xác định độ phân biệt của từng câu hỏi trong bài kiểm tra. Chẳng hạn, đối với câu hỏi 19. Sau khi chấm điểm, trong số 27% tổng số HS (33 em) thuộc nhóm giỏi có 25 em trả lời đúng, đối với nhóm kém có 11 em trả lời đúng câu hỏi này. Như vậy, độ phân biệt của câu hỏi 19 là:

$$D = \frac{C - T}{S} = \frac{25 - 11}{33} = 0,4242$$

Có một số quan niệm khác về độ phân biệt của câu hỏi, bạn đọc có thể tìm thấy trong tài liệu [1], [2]. Thông thường, một câu trắc nghiệm có độ phân biệt lớn hơn 0.2 được xem là chấp nhận được.

Ngoài việc tính độ phân biệt như trên, giáo viên tiến hành thống kê việc lựa chọn các phương án trả lời của hai nhóm thí sinh giỏi và kém như đã nói ở trên. Căn cứ vào thống kê này, giáo viên cũng có thể phát hiện những lỗi sai về kĩ thuật viết câu hỏi cũng như phát hiện các lỗi trong các phương án trả lời đã đưa ra. Chi tiết, bạn đọc có thể tham khảo trong tài liệu [1].

## 3. Căn cứ vào các phần mềm phân tích

### 3.1. Phần mềm CONQUEST

CONQUEST là một phần mềm bản quyền của Hội đồng Australian về nghiên cứu giáo dục (ACER). Đây là phần mềm được sử dụng một cách rộng rãi trên thế giới trong việc phân tích, xử lí số liệu. Có nhiều chức năng trong phần mềm CONQUEST, bạn đọc có thể xem trong tài liệu [8]. Trong bài báo này, chúng tôi giới thiệu các phân tích theo lí thuyết trắc nghiệm cổ điển. Theo cách thống kê này, phần mềm sẽ đếm số lượng học sinh trả lời từng phương án của câu hỏi trắc nghiệm khách quan nhiều phương án lựa chọn, đồng thời chỉ ra độ khó cùng với các thông tin khác về các câu hỏi đó.

Ví dụ, câu hỏi số 6: Cho tập hợp  $M = \{1; 2; 5; 9\}$ . Cách viết nào sau đây là đúng?

1.  $\{2\} \in M$

2.  $9 \subset M$

3.  $\{1; 5; 9\} \subset M$

4.  $\{2; 5\} \in M$

Kết quả thống kê sau khi phân tích như sau:

Câu hỏi: 6 (Toán 6)

Số lượng thí sinh tham gia trả lời câu hỏi  
(Cases for this item): 121

Nhãn (Label)	Điểm (Score)	Đếm (Count)	Tỉ lệ phần trăm (% of tot)
1	0.00	26	21.49
2	0.00	14	11.57
3	1.00	64	52.89
4	0.00	12	9.92
M	0.00	5	4.13

Căn cứ vào thống kê từ bảng trên, có tất cả 121 HS tham gia trả lời câu hỏi này, trong đó có 26 (21.49%) HS lựa chọn phương án 1, 14 (11.57%) HS lựa chọn phương án 2, 64 (52.89%) HS lựa chọn phương án 3 (là phương án đúng), 12 (9.92%) HS lựa chọn phương án 4 và có 5 (4.13%) HS không hoàn thành câu hỏi này.

Sử dụng các thông tin trên đây, chúng tôi xem xét các câu hỏi trong đề kiểm tra đã thử nghiệm. Đa số các câu hỏi nhìn chung là chấp nhận được theo yêu cầu về độ khó, độ phân biệt, cũng như về chất lượng của các câu hỏi trắc nghiệm khách quan. Tuy nhiên, có một số câu hỏi cần được chú ý. Cụ thể là:

a. Câu hỏi 16: Kết quả  $5^{15} : 5^5$  bằng số nào sau đây?

1.  $5^5$
2. 1
3.  $5^3$
- D.  $5^{10}$ .

Kết quả thống kê cho ta bảng sau:

Câu hỏi: 16 (Toán 16)

Số lượng thí sinh tham gia trả lời câu hỏi  
(Cases for this item): 121

Nhãn (Label)	Điểm (Score)	Đếm (Count)	Tỉ lệ phần trăm (% of tot)
1	0.00	21	17.36
2	0.00	9	7.44
3	1.00	46	38.02
4	0.00	42	34.71
M	0.00	3	2.48

Đối với câu hỏi này, đã có đến 46 học sinh (38.02%) lựa chọn phương án 3, trong khi đó, phương án đúng là phương án 4 chỉ có 42(34.71%) HS lựa chọn trả lời. Như vậy, câu hỏi

16 này có vấn đề. Khi đó, từ nội dung câu hỏi, ta dự đoán lí do có thể là HS chưa nắm vững được kiến thức đang kiểm tra. Ngoài ra khi gặp trường hợp tương tự, có thể xem xét một trong các lí do (i) có thể phương án đúng là phương án 3 chứ không phải phương án 4; hoặc (ii) câu hỏi tạo ra sự hiểu lầm cho HS; hoặc (iii) cả hai phương án 3 và 4 đều đúng... Như vậy, tác giả của câu hỏi cần xem xét lại xem nguyên nhân ở đâu và điều chỉnh nó nếu cần thiết.

b. Câu hỏi 25. Cho tập hợp  $A = \{8, 12, 14\}$ . Cách viết nào sau đây không đúng?

1.  $14 \in A$
2.  $\{8, 12, 14\} \subset A$
3.  $\{8, 14\} \subset A$
4.  $\{12\} \in A$ .

Kết quả thống kê như sau:

Câu hỏi: 25 (Toán 25)

Số lượng thí sinh tham gia trả lời câu hỏi  
(Cases for this item): 121

Nhãn (Label)	Điểm (Score)	Đếm (Count)	Tỉ lệ phần trăm (% of tot)
1	0.00	50	41.32
2	0.00	8	6.61
3	1.00	9	7.44
4	0.00	14	11.57
M	0.00	40	33.06

Trong trường hợp này, phương án 1 là đáp án của câu hỏi và có 50 học sinh trả lời đúng, tuy nhiên câu hỏi này cần được xem xét vì có đến 40 HS không hoàn thành việc trả lời. Điều này có thể do các lí do (i) câu dẫn của câu hỏi không rõ ràng; hoặc (ii) phương án đúng không thực sự rõ ràng; hoặc (iii) cách dạy của giáo viên làm cho HS hiểu lầm về mặt kiến thức... Tác giả cũng nên xem xét lại câu hỏi này.

c. Câu hỏi 11. Số nào sau đây chi hết cho cả 3 và 5

1. 25
2. 50
3. 75
4. 18

Thống kê thu được từ câu hỏi này là:

Câu hỏi: 11 (Toán 11)

Số lượng thí sinh tham gia trả lời câu hỏi  
(Cases for this item): 121

Câu hỏi 11 này cũng gặp vấn đề khi phương

Nhãn (Label)	Điểm (Score)	Đếm (Count)	Tỉ lệ phần trăm (% of tot)
1	0.00	22	18.18
2	0.00	84	69.42
3	1.00	13	10.74
4	0.00	1	0.83
M	0.00	1	0.83

án nhiều số 4 chỉ có 01 em HS lựa chọn để trả lời. Lỗi này có khả năng do phương án nhiều số 4 là phương án không nằm trong các sai lầm mà HS thường mắc phải mà là một phương án được cho một cách tùy tiện để có đủ 4 phương án lựa chọn, do đó ngay cả một HS có năng lực thấp vẫn biết đó không thể là phương án trả lời. Đây là một trong những vấn đề giáo viên các trường phổ thông hay mắc phải khi soạn các câu hỏi trắc nghiệm có nhiều phương án lựa chọn. Khi không thể tìm ra được đủ các sai lầm của HS để xây dựng các phương án nhiễu, giáo viên thường lấy phương án bất kì để làm phương án nhiễu. Điều này không nên thực hiện vì rõ ràng phương án như vậy sẽ không có giá trị đo lường, bên cạnh đó, xác suất trả lời đúng câu hỏi đó do đoán mò sẽ tăng lên so với các câu hỏi khác.

### 3.2. Phần mềm BILOG-MG3

BILOG-MG3 là phần mềm phân tích câu hỏi thuộc tập đoàn hệ thống đánh giá (ASC) của Mỹ. Như trên chúng tôi cũng đã đề cập đến, việc thử nghiệm các câu hỏi kiểm tra còn cung cấp các tham số và thông tin về câu hỏi đó. Theo thuyết ứng đáp câu hỏi (IRT), thông qua việc thử nghiệm với một cỡ mẫu học sinh phù hợp, bằng cách sử dụng phần mềm BILOG-MG3 để phân tích số liệu, chúng ta thu được các tham số về độ khó, độ phân biệt, độ phỏng đoán cũng như đường cong thông tin, đường cong đặc trưng của các câu hỏi và các thông tin khác.

Các yếu tố này sẽ giúp cho việc so sánh các bài kiểm tra nhằm đánh giá kết quả học tập của HS một cách chính xác phù hợp với năng lực của HS và cũng từ các yếu tố này sẽ là cơ sở cho việc xây dựng ngân hàng câu hỏi, một nội dung mà chúng tôi sẽ giới thiệu trong dịp khác.

## 4. Kết luận

Thông qua các phân tích trên đây, chúng ta có thể thấy rằng việc thử nghiệm các câu hỏi trước khi triển khai kiểm tra chính thức, đặc biệt đối với các kì thi lớn là hết sức quan trọng. Thông qua thử nghiệm, bằng phương pháp quan sát cũng như phương pháp thông kê, các câu hỏi sẽ được phân tích, xử lý từ đó điều chỉnh các câu hỏi cần thiết nhằm đảm bảo đánh giá mang lại hiệu quả và chất lượng cao. Các phương pháp đã nêu trên đây có thể được áp dụng một cách không khó khăn tại các cơ sở giáo dục vì các phép tính để thực hiện là đơn giản nhưng lại mang đến một kết quả tốt hơn nhiều so với việc ra đề kiểm tra theo lối kinh nghiệm như hiện nay.

### TÀI LIỆU THAM KHẢO

1. Dương Thiệu Tống, *Trắc nghiệm và đo lường thành quả học tập (phương pháp thực hành)*, NXB Khoa học xã hội, 2005.
2. Lâm Quang Thiệp, *Trắc nghiệm và ứng dụng*, Nhà xuất bản Khoa học và Kỹ thuật, 2008.
3. Frank B. Baker, *The Basic of Item Response Theory*, ERIC Clearinghouse on Assessment and Evaluation, 2001.
4. Trần Thị Bích Liễu, *Đánh giá chất lượng giáo dục, nội dung-phương pháp-kỹ thuật*, NXB Đại học Sư phạm, 2007.
5. Nguyễn Phụng Hoàng, Võ Ngọc Lan, *Phương pháp trắc nghiệm trong kiểm tra và đánh giá thành quả học tập*, NXB Giáo dục, 1996.
6. Dự án Phát triển Giáo dục THCS II, *Tài liệu tập huấn đánh giá kết quả học tập của học sinh*, 2006.
7. Dự án Phát triển Giáo dục THCS II, *Tài liệu tập huấn xây dựng ngân hàng câu hỏi*, 2007.
8. Margret L. Wu, Raymond J. Adams, Mark R. Wilson, Samuel A. Haldane, *Acer ConQuest Version 2.0: Generalised Item Response Modelling software*, Australian Council for Educational Research, 2007.

### SUMMARY

*In order to improve education quality through test questions the author proposes methods for checking test question by scientific evidence: difficulty, differentiation... using analysis software. Checking the making of questions will get better results as the common experience practices in making test questions.*