

SỬ DỤNG HIỆU QUẢ CÁC DẠNG THỨC CÂU HỎI THI - KIỂM TRA

• TS. PHẠM XUÂN THANH

Bộ Giáo dục và Đào tạo

Thi - kiểm tra (testing) là một lĩnh vực khoa học có quá trình phát triển lâu dài ở nhiều nước trên thế giới nhưng chưa được quan tâm nhiều ở Việt Nam. Trước đây, nhiều người, nhiều nơi mới chỉ quan tâm đến việc giảng dạy, chưa quan tâm đến thi - kiểm tra. Nhiều người chỉ quen dùng mỗi một dạng đề tự luận và phê phán các dạng đề lựa chọn câu trả lời (trắc nghiệm). Nhưng gần đây, nhiều người lại quá đề cao dạng đề trắc nghiệm, xem đó như một giải pháp mới trong thi - kiểm tra. Hơn nữa, do không được trang bị đầy đủ lí thuyết thi - kiểm tra (test theory) nên trong thời gian qua, một số nơi đã phạm những sai lầm trong việc ra đề, dẫn đến kết quả thi thiếu chính xác, có độ tin cậy kém và thiếu hiệu quả.

Bài báo này nhằm trao đổi một số ý kiến về vai trò của việc thi - kiểm tra, so sánh các dạng câu hỏi làm cơ sở cho việc lựa chọn một dạng thức thích hợp cho mỗi kì thi - kiểm tra.

Các đề thi - kiểm tra là công cụ để đo lường và đánh giá mức độ nhận thức của học sinh (HS). Vậy làm thế nào để đánh giá một cách hiệu quả nhất? Theo Gronlund (1998) và Linn & Gronlund (1995), bộ công cụ để đo lường và đánh giá trong giáo dục ít nhất cũng phải đảm bảo các yêu cầu sau:

- * Có các khái niệm rõ ràng về các kết quả học tập dự định đánh giá (mong muốn HS đạt được những kiến thức và kĩ năng nào, phân bậc các kiến thức và kĩ năng đó, các tiêu chí đánh giá);

- * Đề thi - kiểm tra đáp ứng được mục tiêu của kì thi - kiểm tra

- * Sử dụng các dạng thức thi - kiểm tra khác nhau để loại trừ nhược điểm của từng dạng thức;

- * Dạng thức thi - kiểm tra phải phù hợp với các kết quả học tập dự định đo lường, với các thông tin dự định phản hồi với HS;

- * Có số lượng câu hỏi - bài tập thích hợp để đánh giá đầy đủ và chính xác nhận thức của HS;

- * Quy trình thi - kiểm tra phải công bằng với mọi HS (rõ ràng, không có thiên kiến, phù hợp và được sử dụng đúng chỗ);

- * Có các tiêu chí cụ thể để phân tích, lí giải các kết quả đạt được của HS;

- * Có thông tin phản hồi cho HS, nhấn mạnh những điểm mạnh cần phát huy và điểm yếu cần khắc phục. Thông tin phản hồi cần phải: nhanh chóng; cụ thể; chỉ rõ những điểm mạnh và điểm yếu; cách thức khắc phục; có thiện chí giúp đỡ HS;

- * Kết quả đo lường phải hỗ trợ cho việc chấm điểm (đối với người học) và hỗ trợ hệ thống báo cáo của nhà trường

Đề thi - kiểm tra được cấu thành từ một hay một số câu hỏi hay các yêu cầu (gọi chung là các câu hỏi). Osterlind (1989) đã đưa ra các tiêu chí để đánh giá từng câu hỏi trong mỗi đề thi - kiểm tra:

- * Mục tiêu chính của kì thi - kiểm tra và mục tiêu của từng câu hỏi cần được xác định rõ

- * Mỗi một câu hỏi phải nhằm đạt được mục tiêu chính của kì thi - kiểm tra

- * Các câu hỏi phải đảm bảo sai số trong đo lường ở mức nhỏ nhất (sai số ngẫu nhiên và sai số do định kiến)

- * Dạng thức của mỗi câu hỏi phải phù hợp với mục tiêu của kì thi - kiểm tra

- * Mỗi câu hỏi cần đảm bảo tính đồng nhất (sự ổn định) và tính độc lập cục bộ (độ phân biệt)

- * Các câu hỏi phải được diễn đạt rõ ràng, ngắn gọn, đúng chính tả

- * Mỗi một câu hỏi phải đảm bảo tính hợp pháp và phù hợp với văn hoá, đạo đức

Các nhà giáo dục đã phân chia các câu hỏi thành dạng câu hỏi yêu cầu lựa chọn trả lời (gọi là câu hỏi trắc nghiệm) và dạng câu hỏi yêu cầu trả lời theo cấu trúc (gọi là câu hỏi tự luận). Tuy theo mục đích sử dụng mà các câu hỏi cũng được phân chia thành hai nhóm: để đo lường, đánh giá kiến thức hay kĩ năng. Bảng dưới đây phân loại các câu hỏi theo loại hình và theo mục đích sử dụng (xem bảng 1).

Các dạng câu hỏi này có những ưu, nhược điểm rất khác nhau và chỉ có thể sử dụng có hiệu quả trong từng hoàn cảnh cụ thể. Nói chung, dạng câu hỏi trắc nghiệm có ưu thế để đo lường kiến thức (ví dụ: kiến thức về một môn học) trong quá trình học hay khi kết thúc môn học đó ở các mức nhận thức thấp như nhận biết, hiểu, áp

Bảng 1. Phân loại câu hỏi dùng để đo lường và đánh giá trong giáo dục

Dạng câu hỏi trắc nghiệm	Dạng câu hỏi tự luận
Câu hỏi nhằm mục đích đo lường, đánh giá kiến thức	
<ul style="list-style-type: none"> ▪ Đa lựa chọn ▪ Ghép đôi ▪ Lựa chọn đúng sai ▪ Bộ câu hỏi dựa vào hoàn cảnh cụ thể 	<ul style="list-style-type: none"> ▪ Điền câu trả lời <ul style="list-style-type: none"> ○ Tự luận trả lời ngắn ○ Tự luận trả lời dài ▪ Tự luận miệng ▪ Tự luận làm ở nhà
Câu hỏi nhằm mục đích đo lường, đánh giá kĩ năng	
<ul style="list-style-type: none"> ▪ Bộ câu hỏi dựa vào hoàn cảnh cụ thể 	<ul style="list-style-type: none"> ▪ Câu hỏi yêu cầu thao tác hay trình diễn

dụng... Còn dạng câu hỏi tự luận có ưu thế để đo lường, đánh giá những nhận thức ở mức độ cao (các kĩ năng trình bày, diễn đạt... các khả năng phân tích, tổng hợp, đánh giá,...). Tuy nhiên, với lí thuyết thi - kiểm tra hiện đại, cả hai dạng câu hỏi đều có thể dùng để đo lường, đánh giá những khả năng tư duy ở mức độ cao như giải quyết vấn đề, tư duy sáng tạo hay lí luận phân tích... Ưu nhược điểm của mỗi dạng câu hỏi được xét từ các góc độ: mức độ phức tạp khi viết các câu hỏi và tổ hợp thành đề; tổ chức thi - kiểm tra; sự đoán mò; độ tin cậy và giá trị của đề thi - kiểm tra; chấm điểm; phân tích và đánh giá câu hỏi thi - kiểm tra.

Mức độ phức tạp khi viết câu hỏi và tổ hợp đề

Đứng ở góc độ này, các câu hỏi tự luận có ưu thế hơn các câu hỏi trắc nghiệm vì:

- * Các câu hỏi tự luận dễ làm hơn nhiều so với các câu hỏi trắc nghiệm
- * Các câu hỏi tự luận ít bị sai sót về diễn đạt, có nội dung khá rõ ràng
- * Các câu hỏi trắc nghiệm dễ sai về diễn đạt, nội dung hỏi nhiều khi kém hiệu quả, thiếu chính xác (vì có quá nhiều câu hỏi nên khó rà soát).

Các câu hỏi tự luận bắt buộc phải có đáp án, hướng dẫn chấm, là một việc làm rất phức tạp. Các câu hỏi trắc nghiệm chỉ có 1 phương án đúng cho mỗi câu nên dễ chuẩn bị đáp án và hướng dẫn chấm.

Các đề thi - kiểm tra sử dụng các câu hỏi tự luận khó bảo mật vì đề thi - kiểm tra có ít câu hỏi nên dễ nhớ, dễ lộ. Vì vậy không thể làm đề quá sớm và không thể thử nghiệm trước khi sử dụng nên nhiều khi đề quá dễ hoặc quá khó, không phù hợp với đối tượng dự thi - kiểm tra. Trong các kì thi có quy mô lớn, các đề thi - kiểm tra sử dụng các câu hỏi tự luận ít khi phát huy được hiệu quả cao. Trong các kì thi - kiểm tra giữa môn hoặc kết thúc môn, các câu hỏi tự luận không thể tái sử dụng.

Các đề thi - kiểm tra sử dụng các câu hỏi

trắc nghiệm bao gồm nhiều câu hỏi nhỏ nên khó nhớ, có thể chuẩn bị trước hàng tháng hoặc hàng năm, có thể thử nghiệm trước khi đưa ra sử dụng. Trong các kì thi có quy mô lớn, với kĩ thuật thi - kiểm tra hiện đại và công nghệ thông tin tiên tiến, các đề thi - kiểm tra sử dụng các câu hỏi trắc nghiệm sẽ có nhiều thuận lợi hơn so với các đề thi - kiểm tra sử dụng các câu hỏi tự luận.

Mức độ phức tạp khi tổ chức thi - kiểm tra

Các câu hỏi tự luận đòi hỏi phải dành nhiều thời gian cho HS làm bài vì phải viết câu trả lời. Vì vậy, kì thi - kiểm tra chỉ sử dụng câu hỏi tự luận phải kéo dài hàng giờ nhưng vẫn không thể bao phủ toàn bộ chương trình. Các kì thi - kiểm tra sử dụng các câu hỏi trắc nghiệm chỉ cần 40 - 60 phút là đủ cho một đề gồm 50 - 60 câu bao phủ toàn bộ chương trình, chống được việc học tủ, học lệch và luyện thi.

Các đề thi - kiểm tra sử dụng các câu hỏi tự luận thông thường được trình bày trong một trang giấy nên việc in ấn khá đơn giản và không tốn nhiều thời gian. Các đề thi - kiểm tra sử dụng các câu hỏi trắc nghiệm với số lượng khoảng 50 - 60 câu thường được bố trí trên 4-5 trang giấy nên cần có nhiều thời gian cho việc in ấn. Những rắc rối phức tạp có thể xảy ra khi ghép nhầm các trang của đề thi. Tuy nhiên, việc in các đề thi - kiểm tra sử dụng các câu hỏi trắc nghiệm thường được thực hiện sớm từ trước tại nơi cung cấp đề thi vì thế có thể hạn chế được những sai sót khi nhận đề.

Trong các kì thi - kiểm tra có sử dụng các câu hỏi tự luận, HS dễ dùng phao, tài liệu trong khi làm bài, thậm chí nhìn bài của nhau. Một đề lọt ra ngoài có thể dẫn đến việc nhiều người có bài làm sẵn ở ngoài rồi đưa vào phòng thi.

Với các đề thi - kiểm tra sử dụng các câu hỏi trắc nghiệm theo cách thông thường, HS vẫn có thể quay cốp, xem bài của nhau. Tuy nhiên, đề thi có nhiều câu hỏi thì HS khó có thể sử dụng tài liệu. Các câu hỏi trong đề thi được sắp xếp theo nhiều phương án khác nhau làm cho HS ngồi gần nhau khó có thể chép bài của nhau; bài

làm ở ngoài đưa vào chỉ có một người có thể sử dụng được; nếu có hiện tượng tiêu cực trong phòng thi thì không ảnh hưởng đến nhiều người. Nếu sử dụng đề có 50-60 câu hỏi đa lựa chọn, các câu hỏi được bố trí theo nhiều phương án khác nhau, các câu hỏi giống nhau có các phương án trả lời được bố trí khác nhau, có thể loại trừ hầu hết tiêu cực của HS.

Như vậy, đứng về mặt tổ chức thi – kiểm tra, trong các kì thi có quy mô lớn, các đề thi - kiểm tra sử dụng các câu hỏi trắc nghiệm sẽ làm đơn giản hoá quá trình coi thi, chống được tiêu cực trong phòng thi.

Vấn đề đoán mò phương án trả lời đúng

Với các câu hỏi tự luận, HS không có cơ hội đoán được câu trả lời đúng. Nhưng với các câu hỏi trắc nghiệm, HS có thể đoán được phương án trả lời đúng. Đây là một nhược điểm lớn, nhất là đối với những câu hỏi có chất lượng thấp. Với một câu hỏi có bốn phương án lựa chọn, xác suất đoán đúng phương án trả lời là 0.25. Nhưng với các câu hỏi có chất lượng thấp, xác suất đoán đúng phương án trả lời còn cao hơn vì HS có thể loại trừ những phương án sai quá lộ liễu hoặc nhìn thấy phương án đúng quá rõ ràng. Tuy nhiên, với một đề thi có nhiều câu hỏi trắc nghiệm thì xác suất đoán đúng nhiều câu là rất thấp. Xác suất trả lời đúng cả 10 câu hỏi trắc nghiệm (4 lựa chọn) có chất lượng tốt là 0.000000009. Với đề thi có trên 50 câu hỏi trắc nghiệm thì kết quả trả lời đúng 20-30% số câu hỏi chủ yếu là do HS đoán mò.

Nếu một đề thi có 60 câu có 4 lựa chọn và một HS trả lời hết tất cả các câu, trong đó có 50 câu đúng, khi đó số câu trả lời đúng bằng thực lực của HS là 47 câu hoặc hơn và số câu đúng nhờ đoán mò, tối đa là 3 câu.

Vi vậy, để giảm thiểu ảnh hưởng của việc đoán mò đến việc đánh giá chính xác thực lực của HS, mỗi đề thi trắc nghiệm nên có khoảng 50 – 60 câu và mỗi câu nên có 4 phương án lựa chọn.

Độ tin cậy của đề thi – kiểm tra

Độ tin cậy là điều kiện cần của các đề thi – kiểm tra để đánh giá chính xác kết quả học tập của HS: HS có năng lực tốt hơn phải đạt được kết quả cao hơn. Độ tin cậy có thể tăng lên nhờ bổ sung những câu hỏi đồng nhất (tăng độ dài của đề thi) hay tăng độ phân biệt của các câu hỏi.

Đề trắc nghiệm có độ tin cậy cao hơn đề tự luận vì giảm được tính chủ quan khi chấm.

Tính giá trị của đề thi – kiểm tra

Đề thi – kiểm tra có giá trị cao khi giúp chúng ta đạt được mục đích của kì thi – kiểm tra. Đề thi – kiểm tra phải đảm bảo công bằng và đại

diện cho nội dung cần đánh giá. Đề tự luận làm trong một giờ để đánh giá kết quả học tập một học kì của HS không thể bao quát toàn bộ chương trình nên có thể có những sai lệch không nhỏ. Đề thi – kiểm tra gồm nhiều câu hỏi tự luận ngắn hay dạng điền khoảng trống có thể giúp để bao quát chương trình ở mức độ rộng hơn. Đề trắc nghiệm 50 – 60 câu làm trong một tiếng có thể bao quát toàn bộ chương trình và cho phép đánh giá chính xác hơn đề tự luận.

Đề thi – kiểm tra phải có khả năng đánh giá HS theo các mức nhận thức từ thấp đến cao. Trái với định kiến của nhiều người, đề trắc nghiệm cũng như đề tự luận đều có thể dùng để đánh giá các mức nhận thức từ mức thấp (nhận biết, hiểu) đến các mức cao như phân tích, tổng hợp, đánh giá. Kết quả chấm điểm của hai dạng thức đề thi cho cùng một nội dung cần đánh giá có một mối tương quan chặt chẽ với nhau, vì thế cả hai dạng thức có thể sử dụng thay thế nhau.

Sự phức tạp của việc chấm điểm

Chấm điểm là một việc làm khá phức tạp, gây nhiều tranh cãi vì có thể phụ thuộc nhiều vào người chấm. Đề tự luận thường sử dụng phương pháp chấm điểm theo đáp án hoặc theo thang ước lượng nên vẫn phụ thuộc nhiều vào người chấm. Cùng một bài làm nhưng những người chấm khác nhau hoặc các trường chấm khác nhau có thể cho kết quả khác nhau. Việc cho điểm các đề tự luận có thể bị ảnh hưởng bởi chữ viết của HS, độ dài của câu văn, phong cách viết, giới tính...

Trong một lớp học, tính chủ quan của giáo viên không ảnh hưởng nhiều đến sự công bằng khi chấm các bài làm của HS. Nhưng trong các kì thi có quy mô lớn, điều đó có thể dẫn đến việc thiếu công bằng nghiêm trọng. Đặc biệt, trong các kì tuyển sinh, tính thiếu khách quan trong chấm thi có thể làm cho nhiều HS bị thiệt thòi. Trong khi đó, các đề thi trắc nghiệm có thể chấm bằng cách sử dụng khoá (key), phiếu chấm hay máy quét quang học nên rất khách quan và đảm bảo công bằng cho HS; đồng thời có thể giảm được số người tham gia chấm, hạn chế được tiêu cực và rút ngắn thời gian chấm.

Đứng trên góc độ chấm điểm các bài thi trong các kì thi quy mô lớn thì các đề trắc nghiệm có ưu thế hơn đề tự luận.

Mức độ phức tạp khi phân tích và đánh giá các câu hỏi thi – kiểm tra

Phân tích và đánh giá các câu hỏi thi – kiểm tra là một việc rất cần thiết nhưng còn ít được quan tâm ở nhiều nơi. Hậu quả là các kì thi sau vẫn thường lặp lại những sai lầm của các kì thi trước đó. Việc phân tích, đánh giá các câu hỏi sẽ giúp chúng ta hiểu rõ hơn chất lượng của đề thi

và chất lượng của HS. Tuy nhiên, mức độ phức tạp của việc phân tích, đánh giá các câu hỏi phụ thuộc nhiều vào dạng thức câu hỏi được sử dụng trong kì thi.

Thực tế cho thấy rất khó phân tích và đánh giá các câu hỏi tự luận. Rất khó đánh giá độ khó của các câu hỏi này, khó xác định những câu tự luận mơ hồ để loại bỏ hoặc sửa chữa chúng. Trong khi đó, với các câu hỏi trắc nghiệm đã có khá nhiều phần mềm để phân tích đánh giá. Nhờ đó, các câu hỏi trắc nghiệm thường xuyên được nâng cao chất lượng, loại bỏ hoặc sửa chữa được những câu thiếu rõ ràng, nâng được độ tin cậy và tính giá trị của các đề thi – kiểm tra.

Do không chú ý đến việc phân tích đề thi, không có thử nghiệm trước khi thi nên một số đề thi – kiểm tra quá dễ hoặc quá khó so với năng lực của HS. Đề quá dễ dẫn đến kết quả làm bài của HS quá tốt nên bị xã hội phê phán là chạy theo thành tích, không phản ánh đúng thực lực của HS. Đề quá khó dẫn đến kết quả làm bài của HS quá kém nên bị xã hội phê phán là chất lượng giáo dục thấp. Tuy nhiên, đánh giá chất lượng giáo dục mà chỉ nhìn vào điểm số là cách đánh giá rất thiếu chính xác vì điểm số phụ thuộc rất nhiều vào độ khó của đề thi và tính chủ quan của người chấm. Không thể so sánh kết quả hai kì thi khác nhau khi các đề thi đó có độ khó khác nhau. Thậm chí, cùng một đề thi nhưng do những người chấm khác nhau cũng dẫn đến kết quả khác nhau. Gần đây một số tác giả dùng các biểu đồ phân bố kết quả điểm của HS để mô tả chất lượng giáo dục là thiếu thuyết phục, bởi các biểu đồ kết quả phụ thuộc vào độ khó của đề.

Các nhà giáo, các nhà quản lí giáo dục cần thống nhất rằng mục đích cho điểm trong các kì thi – kiểm tra chỉ là thứ yếu; mục đích chính là cung cấp thông tin chính xác về thực trạng chất lượng giáo dục: HS có những khả năng gì và năng lực của HS khác nhau như thế nào giữa các trường, các tỉnh, các khu vực...

Đành rằng, ngày nay, các giáo viên, các trường học và cả các tổ chức giáo dục ở nhiều nước trên thế giới đang phải chịu những sức ép ghê gớm từ phía HS, phụ huynh và xã hội trong việc chấm điểm cho HS. Họ phải chấm cho HS ở mức điểm cao hơn mức HS thực sự đạt được. Chính vì lí do đó mà các nước phải tổ chức các kì thi quốc gia (không nhất thiết là kì thi cuối cấp) để đánh giá đúng thực trạng giáo dục và để ra các giải pháp kịp thời. Để đạt được độ khách quan và chính xác cao, các nước thường dùng các đề thi đã được chuẩn hoá (standardised tests), đã được thử nghiệm và đã được xác định độ khó. Điều quan trọng của các kì thi này không

phải để chấm điểm HS mà là để chỉ ra HS có đạt được các chuẩn kiến thức theo thiết kế hay không.

Kết luận

Những phân tích, so sánh ở trên cho thấy:

- Mỗi dạng thức câu hỏi đều có ưu nhược điểm riêng tùy theo trường hợp sử dụng cụ thể, vì vậy cần cân nhắc lựa chọn dạng thức thích hợp trong mỗi kì thi – kiểm tra để đạt được hiệu quả sử dụng cao nhất.

- Đề tự luận nên sử dụng để đánh giá kết quả học tập trong phạm vi nhỏ (do ít người chấm), khi không có nhiều thời gian để làm các câu hỏi trắc nghiệm và khi cần đánh giá khả năng diễn đạt của HS.

- Các đề trắc nghiệm đã được chuẩn hoá có thể sử dụng để đánh giá kiến thức của HS trong quá trình giảng dạy các môn học.

- Trong các kì thi có quy mô lớn nên sử dụng đề thi trắc nghiệm để chống tiêu cực trong khi làm bài, đảm bảo tính khách quan, công bằng khi chấm. Các câu hỏi trắc nghiệm cần được chuẩn bị kĩ, được thử nghiệm trước khi sử dụng, nhất là đối với những câu hỏi nhằm đánh giá khả năng tư duy của HS ở mức độ cao.

- Không thể so sánh kết quả học tập của HS nếu sử dụng các đề thi – kiểm tra có độ khó khác nhau. Muốn so sánh, nhất thiết phải sử dụng một đề thi chung hoặc sử dụng những câu hỏi chung để kết nối chúng với nhau.

- Nhằm hỗ trợ cho các giáo viên giảng dạy tốt hơn, cần thiết phải đưa các kiến thức về đo lường và đánh giá trong giáo dục vào các chương trình đào tạo ở các trường sư phạm và đưa vào các chương trình tập huấn hàng năm.

TÀI LIỆU THAM KHẢO

1. Gronlund, N. E. (1998). *Assessment of student achievement*. Boston: Allyn and Bacon.
2. Haladyna, T.M. (1994). *Developing and Validating Multiple Choice Test Items*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
3. Hambleton, R.K. & Swaminathan, H. (1985). *Item Response Theory, Principles and Applications*. Kluwer, Nijhoff Publishing, Boston.
4. Popham, W. J. (1998). *Classroom assessment: What teachers need to know*. Boston: Allyn and Bacon.

SUMMARY

The article presents, compares and analyzes various forms of questions in tests-examinations on which basis to choose the most suitable and effective form for teach test or examination.