

NGHIÊN CỨU VỀ NGÂN HÀNG CÂU HỎI ĐÁNH GIÁ NĂNG LỰC THÍ SINH

ĐỖ ĐÌNH THÁI - Email: dodinhthai@sgu.edu.vn

LÊ CHI LAN - Email: chilanih@yahoo.com

Trường Đại học Sài Gòn

Tóm tắt: Ngân hàng câu hỏi đóng vai trò quan trọng trong quá trình đánh giá chất lượng đào tạo về kiến thức, kỹ năng, thái độ của người học và đánh giá năng lực của thí sinh trong một kì thi. Để đánh giá xác thực năng lực của thí sinh thì ngân hàng câu hỏi có chất lượng là thành phần không thể thiếu. Dựa trên cơ sở phân tích lợi ích của ngân hàng câu hỏi, những hạn chế đối với ngân hàng câu hỏi, bài viết nghiên cứu một số phương pháp định cỡ câu hỏi, cách thức xây dựng và triển khai phương pháp đánh giá năng lực thích ứng thông qua ngân hàng câu hỏi đã được định cỡ các câu hỏi để đánh giá, phân loại năng lực thí sinh trong các kì thi. Trong đó, phương pháp định cỡ câu hỏi quyết định chất lượng ngân hàng câu hỏi, phương pháp đánh giá năng lực thích ứng nhằm tối ưu hóa thời gian, cách tổ chức đánh giá mà vẫn đảm bảo phân loại được năng lực thí sinh. Qua cách tiếp cận trên, các trường đại học có thể nghiên cứu áp dụng trong việc kiểm tra, đánh giá năng lực thí sinh ở các kì thi.

Từ khóa: Ngân hàng câu hỏi; đánh giá; năng lực; sinh viên.

(Nhận bài ngày 16/3/2017; Nhận kết quả phản biện và chỉnh sửa ngày 15/4/2017; Duyệt đăng ngày 25/6/2017).

1. Đặt vấn đề

Ngân hàng câu hỏi đóng vai trò quan trọng trong quá trình đánh giá chất lượng đào tạo về kiến thức, kỹ năng và thái độ của người học đồng thời đánh giá năng lực của thí sinh trong một kì thi. Tính khách quan của ngân hàng câu hỏi càng cao thì minh chứng cho năng lực của người học càng được thể hiện rõ nét, đảm bảo công bằng trong phân loại năng lực thí sinh của một kì thi. Do vậy, các yếu tố ảnh hưởng đến chất lượng ngân hàng câu hỏi cần được quan tâm, đầu tư đúng mức để đảm bảo đáp ứng mục tiêu môn học, chuẩn đầu ra của chương trình đào tạo, nhu cầu thị trường lao động, thỏa mãn các điều kiện kiểm tra, đánh giá năng lực thí sinh. Để thi được tạo ra từ ngân hàng câu hỏi phải đảm bảo độ tin cậy và độ giá trị tốt, đánh giá, phân loại được năng lực thí sinh. Đặc biệt, ngân hàng câu hỏi phải đánh giá được các năng lực khác nhau của người học trong quá trình đào tạo là một trong những điều kiện cốt lõi hình thành sản phẩm chất lượng của một cơ sở giáo dục.

2. Ngân hàng câu hỏi

Hiện nay, ngân hàng câu hỏi theo cụm từ tiếng Anh là "item bank" hoặc "question bank". Tuy nhiên, cộng đồng đánh giá quốc tế nói chung thường sử dụng thuật ngữ "item bank" hơn vì nó mang ý nghĩa cụ thể trong từng bối cảnh đánh giá trong khi "question bank" mang giá trị ngữ nghĩa (nội hàm) cụ thể đối với một số dạng đánh giá [1].

Ngân hàng câu hỏi là một kho lưu trữ các câu hỏi kiểm tra và các thành phần tạo nên câu hỏi. Ngân hàng câu hỏi được biên soạn tốt và quản lý cẩn thận có thể tạo ra các bài kiểm tra đo lường chính xác năng lực về kiến

thức, kỹ năng và khả năng cần thiết của thí sinh trong các kì thi [2], là một tập các câu hỏi được xây dựng để đo lường phạm vi kiến thức hoặc kỹ năng được xác định cụ thể. Ngân hàng câu hỏi chứa các đặc trưng khác nhau ở mỗi câu hỏi. Các đặc trưng này có thể liên quan đến nội dung hoặc quản lý thông tin như các đặc trưng về mức độ nhận thức, các tham số câu hỏi [3]. Ngân hàng câu hỏi bao gồm việc lưu trữ các câu hỏi và thông tin về câu hỏi (như độ khó, độ phân biệt, độ tin cậy, độ hiệu lực,...) dưới hình thức điện tử [4].

Ngân hàng câu hỏi trắc nghiệm (test bank) là tài nguyên trắc nghiệm được tạo ra trên máy tính. Giáo viên có thể hiệu chỉnh, bổ sung cho phù hợp với mục đích giảng dạy của họ. Dạng câu hỏi trắc nghiệm bao gồm: nhiều lựa chọn, trả lời một hoặc nhiều phương án đúng, đúng sai, điền khuyết, ghép đôi, trả lời ngắn và tiểu luận [5]. Câu hỏi trắc nghiệm được chia thành 2 nhóm trắc nghiệm khách quan (nhiều lựa chọn, đúng sai, điền khuyết, ghép đôi và trả lời ngắn) và trắc nghiệm tự luận (tiểu luận và cung cấp thông tin) [4], trong đó mỗi câu hỏi đã được định cỡ, tức là được gắn với các phần nội dung xác định và các tham số xác định (độ khó, độ phân biệt theo lý thuyết trắc nghiệm cổ điển) và các tham số đặc trưng (a, b và c theo lý thuyết ứng đáp câu hỏi) [6].

Diễn hình ngân hàng câu hỏi đánh giá theo dạng thức của ETS (Viện Khảo thí Giáo dục Hoa Kỳ) gồm các câu hỏi có chất lượng cao vì nó luôn thỏa mãn các tiêu chí sau: 1/ Được cập nhật mới nhất; 2/ Được xây dựng và chỉnh lí phù hợp với các bang khác nhau; 3/ Điều chỉnh độ khó và đo lường phạm vi các mức độ nhận thức khác nhau; 4/ Tuân thủ các nguyên tắc tạo ra câu hỏi tốt, dựa



trên các tiêu chuẩn kĩ nghệ; 5/ Sử dụng từ vựng và cấu trúc câu phù hợp với từng lứa tuổi; 6/ Được biên soạn trong ngữ cảnh hợp lí và thực tế; 7/ Sử dụng các tùy chọn đáp ứng được tất cả các tình huống; 8/ Cung cấp các từ gốc trong câu hỏi rõ ràng và dễ hiểu; 9/ Được biên soạn bởi đội ngũ chuyên gia giàu kinh nghiệm của ETS. Ngoài ra, trong hơn 50 năm qua, ETS đã thiết lập các tiêu chuẩn đo lường, nghiên cứu giáo dục, đã phát triển một loạt các hướng dẫn và tiêu chuẩn để đảm bảo chất lượng các câu hỏi. Điều này rất quan trọng đối với một bài kiểm tra công bằng, hợp lí và đáng tin cậy, đảm bảo rằng chỉ các yếu tố có liên quan đến xây dựng câu hỏi ảnh hưởng đến điểm thí sinh [3].

2.1. Lợi ích của ngân hàng câu hỏi

Lợi thế của ngân hàng câu hỏi là tính linh hoạt, tính bảo mật và tính thống nhất cao. Rudner (1998) chỉ ra rằng trong ngân hàng câu hỏi, các câu hỏi có thể được chỉnh sửa, trích ra và lưu trữ thành các câu hỏi mới hoặc các ngân hàng câu hỏi con (theo chủ đề, năng lực,...) khi cần thiết. Ngoài ra, các câu hỏi có thể sử dụng để chẩn đoán bài làm của sinh viên làm tốt hoặc không làm tốt bởi vì các câu hỏi được xác định trên mô hình Rasch cho phép giảng viên xác định độ khó của các câu hỏi dành cho sinh viên [7]. Tương tự như vậy, đối với ngân hàng câu hỏi trắc nghiệm, các câu hỏi trắc nghiệm được cung cấp thông qua hệ thống thích hợp của tổ chức. Giảng viên có thể tiết kiệm được thời gian trong việc tạo ra bài kiểm tra. Giảng viên có thể hiệu chỉnh từng phần trong bài kiểm tra phù hợp với yêu cầu giảng dạy. Sinh viên có thể tham gia tự học thông qua ngân hàng câu hỏi trong một phạm vi nhất định. Giảng viên có thể tích hợp các bài kiểm tra và công cụ kiểm tra đánh giá của họ [5]. Ngân hàng câu hỏi là công cụ hỗ trợ đổi mới phương pháp giảng dạy, kiểm tra đánh giá và quản lí chất lượng đào tạo. Cụ thể, trong giảng dạy, giảng viên thực hiện phương pháp giảng dạy tích cực theo cách tiếp cận năng lực, nâng cao trách nhiệm, cập nhật và bổ sung kiến thức trong quá trình giảng dạy. Trong học tập, giúp người học xác định kiến thức chuẩn của môn học cần phải đạt được, giúp người học tự học và tự kiểm tra nhận thức của bản thân, mở rộng kiến thức qua các tài liệu tham khảo khác nhau, giúp người học hợp tác, trao đổi thảo luận trên cơ sở một số câu hỏi từ ngân hàng câu hỏi.

2.2. Thách thức đối với ngân hàng câu hỏi

Phát triển ngân hàng câu hỏi đưa vào máy tính để đánh giá, phân loại được năng lực thí sinh theo phương pháp đánh giá năng lực thích ứng là một thách thức đối với các nhà chuyên môn và chuyên gia máy tính. Trước tiên, họ phải mất nhiều thời gian để lập kế hoạch và đưa các câu hỏi vào ngân hàng câu hỏi, định cỡ và sau đó đưa các câu hỏi vào cơ sở dữ liệu. Quá trình này đòi hỏi sự kiên nhẫn của các nhà chuyên môn phối hợp với nhiều chuyên gia máy tính. Trong quá trình triển khai ứng

dụng, các câu hỏi phải được tiếp tục đánh giá và định cỡ nhằm đảm bảo tính chính xác cao nhất của độ khó và độ phân biệt của từng câu hỏi [7].

Một bài kiểm tra được đánh giá là có chất lượng tốt khi mỗi câu hỏi được chọn đều có chất lượng tốt. Nếu các câu hỏi không thực sự đo lường được chuẩn đã xác định thì kết quả của các bài kiểm tra sẽ không có giá trị. Do đó, việc định cỡ cho các câu hỏi để đưa vào ngân hàng câu hỏi là yếu tố quan trọng quyết định chất lượng ngân hàng câu hỏi và đánh giá đúng với năng lực của thí sinh.

Hiện nay, các trường đại học thường sử dụng bài kiểm tra trắc nghiệm có chiều dài cố định theo lối truyền thống để đo lường năng lực của sinh viên, nghĩa là toàn bộ sinh viên của một môn học hoặc một kì thi làm một bài kiểm tra trắc nghiệm với số lượng câu hỏi được ước lượng ứng với thời gian quy định. Nhược điểm chính của phương pháp này là khi năng lực của sinh viên vượt xa độ khó của bài trắc nghiệm thì điểm số của sinh viên ấy không được đo chính xác. Ngoài ra, phân bố điểm cho từng câu hỏi chưa phù hợp với độ khó đã được định cỡ dẫn đến kết quả đánh giá chưa đúng với năng lực của sinh viên.

Trong bài trắc nghiệm năng lực thích ứng, bằng cách áp dụng các câu hỏi với độ khó thích hợp ứng với mỗi thí sinh, sai số chuẩn của phép đo sẽ giảm. Nghĩa là, mỗi câu hỏi tiếp theo của mỗi thí sinh được máy tính đưa ra dựa trên kết quả đạt được ở thời điểm hiện tại. Quá trình phát sinh câu hỏi được thực hiện đến khi hoàn tất bài kiểm tra hoặc thí sinh đạt đến mức năng lực tối ưu nhất. Một trong những thành phần chính của các bài trắc nghiệm năng lực thích ứng là ngân hàng câu hỏi gồm các câu hỏi đã được định cỡ.

2.3. Ước lượng số câu hỏi cho ngân hàng câu hỏi

Dựa trên quan điểm *Min - Max: Num* của Sands và các cộng sự (1997), trong đó *Min* là số lượng câu hỏi tối thiểu, *Max* là số lượng câu hỏi tối đa và *Num* là số lượng thí sinh dự thi (ví dụ: 5-10:1, 1 thí sinh dự thi được yêu cầu 5 đến 10 câu hỏi), Burghof (2001) đưa ra công thức tính tỉ lệ ngân hàng câu hỏi = *Min*-*Max*:*Num* như sau:

Tỉ lệ ngân hàng câu hỏi = **Min** (Min x Num) - **Max** (Max x Num) : Num

Giả sử giá trị *Min* = 5, *Max* = 10 và *Num* = 25, khi đó giá trị **Min** là 125 câu hỏi, **Max** là 250 câu hỏi, nghĩa là 25 thí sinh dự thi cần từ 125 đến 250 câu hỏi. Như vậy:

Tỉ lệ ngân hàng câu hỏi = **Min** (125) - **Max** (250) : Num (25) [8].

Đối với số lượng câu hỏi trong ngân hàng câu hỏi trắc nghiệm của một môn học, chúng ta có thể quy định số lượng câu hỏi tối thiểu = số lượng câu hỏi x số tín chỉ môn học. Ví dụ, 150 câu hỏi/1 tín chỉ, môn học X có 3 tín chỉ = 150 câu hỏi x 3 tín chỉ = 450 câu hỏi, số lượng câu hỏi tối đa = số lượng câu hỏi tối thiểu x 2.

Ngoài ra, số lượng câu hỏi trong một đề thi và chu kì bổ sung câu hỏi vào ngân hàng câu hỏi hoặc chu kì thay thế ngân hàng câu hỏi cũng quyết định số lượng câu hỏi tối thiểu và tối đa trong ngân hàng câu hỏi. Khi xác định số lượng câu hỏi, chúng ta cần lưu ý đến thời hạn sử dụng của ngân hàng câu hỏi đối với một số môn học phụ thuộc vào hoặc liên quan đến các văn bản pháp lí, sự phát triển của xã hội, sự phát triển của khoa học ứng dụng,...

3. Phương pháp định cỡ

Định cỡ câu hỏi là yếu tố không thể thiếu trong ngân hàng câu hỏi, nó giúp xác định độ khó và độ phân biệt của từng câu hỏi. Do vậy, việc định cỡ câu hỏi mất nhiều thời gian, công sức và là quá trình thử nghiệm, tích lũy lâu dài để đạt được hiệu quả cao nhất trong các kì thi.

Đối với lí thuyết trắc nghiệm cổ điển, độ khó của câu hỏi bằng tỉ số phần trăm thí sinh làm đúng câu hỏi trên tổng số thí sinh tham gia làm câu hỏi đó, độ phân biệt của câu hỏi bằng $(C - T)/S$, trong đó C là số thí sinh làm đúng câu hỏi thuộc nhóm giỏi (27% thí sinh đạt điểm cao từ trên xuống), T là số thí sinh làm đúng câu hỏi thuộc nhóm kém (27% thí sinh đạt điểm thấp từ dưới lên), S là số thí sinh của một trong hai nhóm [4].

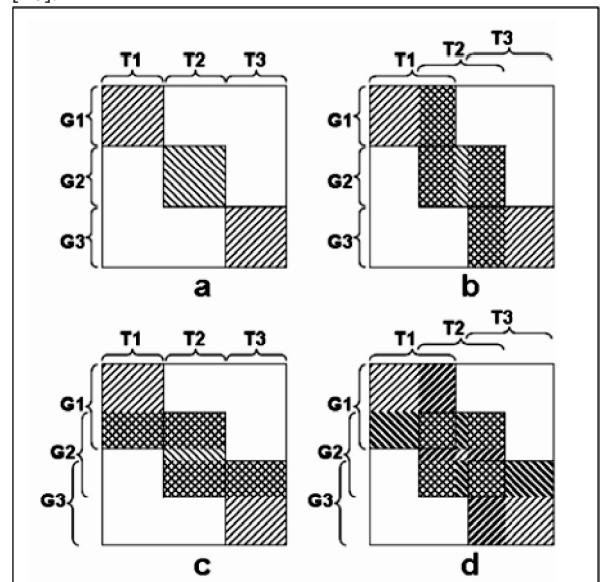
Phương pháp định cỡ của López-Cuadrado và các cộng sự (2008):

- *Định cỡ dựa trên chuyên gia*: gồm các giá trị ước lượng tham số đặc trưng câu hỏi. Khi các quá trình định cỡ và biên soạn được thực hiện riêng rẽ, người ta thường tham khảo ý kiến và quan điểm của một hoặc nhiều chuyên gia cùng lĩnh vực để ước lượng các tham số câu hỏi. Khi sử dụng kinh nghiệm của các chuyên gia, người ta khuyến cáo chỉ ước lượng độ khó của câu hỏi, bởi vì tham số độ phân biệt và hệ số đoán mò có thể không được dự đoán tốt hoặc không được xác định rõ ràng trong đo lường. Trong trường hợp này, người ta có thể sử dụng mô hình Rasch để xác định các tham số này.

- *Định cỡ dựa trên nhận thức*: cho phép đạt được không chỉ độ khó của mỗi câu hỏi trong thang đo sử dụng lí thuyết ứng đáp câu hỏi mà còn độ phân biệt và hệ số đoán mò. Đầu tiên, chúng ta thu thập các phương án trả lời từ bài kiểm tra của một nhóm thí sinh mang tính đại diện (nhóm này cũng là nhóm sẽ tham gia thực hiện các bài kiểm tra sử dụng ngân hàng câu hỏi đã được định cỡ). Do tính chất bảo mật, các câu hỏi được phân loại thành một vài dạng thức kiểm tra khác nhau và được áp dụng riêng. Việc phân loại dạng thức kiểm tra với nhóm thí sinh được quản lí độc lập và không có bất kì mối quan hệ nào. Do đó, giá trị của các ước lượng tham số câu hỏi không chia sẻ thang đo chung. Vấn đề này được giải quyết qua thiết kế câu hỏi neo. Trong các nhóm khác nhau, mỗi thí sinh (không nhất thiết phải tương đương) thực hiện một dạng thức kiểm tra khác nhau nhưng có một vài câu hỏi chung với các nhóm

khác. Sau đó, các ước lượng của các câu hỏi chung (câu hỏi neo) được so sánh, cung cấp giải pháp cân bằng các thang đo của các dạng thức kiểm tra khác nhau để đưa ra được thang đo chung cho các ước lượng tham số của toàn bộ ngân hàng câu hỏi [9].

Molina và các cộng sự (2008) giới thiệu bốn thiết kế định cỡ cơ bản (Hình 1) được xem xét chỉ với 2 ngân hàng câu hỏi con và 2 nhóm đối tượng. Thiết kế (a) minh họa thiết kế định cỡ với 2 bài kiểm tra (T1 và T2) có nội dung không chồng lấn được thực hiện bởi 2 nhóm mẫu (G1 và G2) không chồng lấn. Trong trường hợp này, bằng cách làm cho hoặc 2 bài kiểm tra hoặc 2 nhóm đối tượng tương đương để đạt được liên kết của các tham số câu hỏi trên cùng thang đo. Thiết kế (b) mô tả thiết kế câu hỏi neo, hai nhóm đối tượng (G1 và G2) làm chung một vài câu hỏi (có trong T1 và có trong T2), các câu hỏi neo này là giải pháp chủ yếu cho liên kết các tham số câu hỏi đạt được trên cùng thang đo từ các nhóm đối tượng khác nhau. Thiết kế (c) minh họa thiết kế đối tượng neo, một nhóm đối tượng chung (có trong G1 và có trong G2) thực hiện 2 bài kiểm tra (T1 và T2) và cung cấp giải pháp cho liên kết các tham số câu hỏi trên cùng thang đo. Cuối cùng, thiết kế (d) minh họa một thiết kế neo đôi bao gồm cả các câu hỏi chung và các đối tượng chung [10].



Hình 1: Bốn thiết kế định cỡ

4. Phương pháp kiểm tra đánh giá sử dụng ngân hàng câu hỏi

Hiện nay, đối với ngân hàng câu hỏi trắc nghiệm, phương pháp đánh giá năng lực thích ứng đang được sử dụng rộng rãi nhằm mục đích đánh giá xác thực năng lực của thí sinh trong các kì thi, đặc biệt là các kì thi tuyển sinh vào các trường đại học, cao đẳng. Hiện nay, ở các nước tiên tiến, các phương pháp kiểm tra đánh giá chủ



yếu thực hiện các bài kiểm tra trắc nghiệm trên máy tính để đánh giá và phân loại năng lực thí sinh trong quá trình thí sinh làm bài trong khi các bài kiểm tra trắc nghiệm trả lời trên phiếu hỏi không thực hiện được việc này. Trong quá trình thực hiện bài kiểm tra, thí sinh được đánh giá năng lực liên tục qua từng câu hỏi, kết quả trong quá trình làm bài đến câu hỏi thứ n là ước lượng năng lực đầu vào cho máy tính chọn câu hỏi n + 1 tiếp theo. Cứ như vậy đến khi thí sinh hoàn thành bài kiểm tra hoặc đạt được mức năng lực tốt nhất ứng với năng lực của thí sinh và phù hợp với tiêu chí đánh giá.

Aminifar và Alipour (2014) tổng hợp ý kiến của một số tác giả cho rằng một bài kiểm tra thích ứng gồm 6 thành phần chính:

Lý thuyết đo lường: hệ thống các giả thuyết và suy luận được kết hợp xây dựng để đo lường (Sympson, 1970). Hai lý thuyết chính trong kiểm tra thích ứng là lý thuyết trắc nghiệm cổ điển và lý thuyết ứng đáp câu hỏi (Weiss, 2011).

Ngân hàng câu hỏi: Một bộ câu hỏi có các tham số độ khó, độ phân biệt và độ đoán mò đã được định cỡ (Thompson và Weiss, 2011).

Điểm khởi đầu (năng lực khởi điểm): Trước khi bắt đầu kiểm tra, dựa trên thông tin về năng lực của thí sinh (quá trình học tập, điểm số, kết quả từ bài kiểm tra chung trước,...), một giá trị phù hợp với năng lực khởi điểm của thí sinh được cho trước. Câu hỏi đầu tiên của thí sinh phụ thuộc vào năng lực khởi điểm họ (Weiss, 2004).

Thuật toán lựa chọn câu hỏi: Thuật toán này chọn câu hỏi kế tiếp của bài kiểm tra từ ngân hàng câu hỏi dựa trên năng lực hiện tại của thí sinh (van der Linden, 2005).

Thuật toán tính điểm: Sau khi trả lời một câu hỏi, thuật toán tính điểm ước lượng năng lực mới của thí sinh theo phương án trả lời (Weiss, 2011).

Điều kiện kết thúc bài kiểm tra: Khi các điều kiện liên quan được đáp ứng, quá trình kiểm tra kết thúc (Babcock and Weiss, 2009) [11].

Fong và các cộng sự (2006) giới thiệu phương pháp đánh giá năng lực thích ứng (adaptive test) trên web, là hệ thống đánh giá trực tuyến có thể tạo ra những câu hỏi thích ứng với năng lực sinh viên. Các sinh viên thực hiện một bài kiểm tra với các câu hỏi có độ khó khác nhau (độ khó câu hỏi từ 1 đến 10). Kết quả đạt được của mỗi sinh viên được sử dụng như mức năng lực ước lượng khởi điểm. Mức năng lực tiếp theo sau mỗi câu hỏi được tính dựa trên 2 công thức sau:

$$\text{Giá_trị_mới} = \text{Giá_trị_cũ} + \text{Giá_trị_cũ} * (\text{độ_khó_câu_hỏi} / 10) \quad (1)$$

$$\text{Giá_trị_mới} = \text{Giá_trị_cũ} + \text{Giá_trị_cũ} * ((5 - \text{độ_khó_câu_hỏi}) / 10) \quad (2)$$

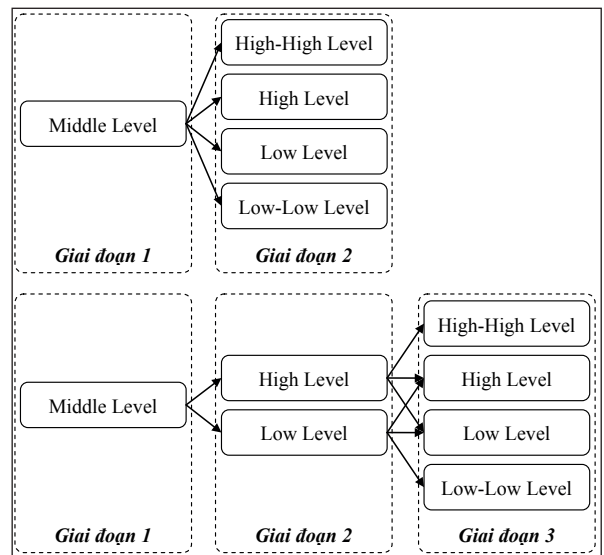
Công thức (1) áp dụng cho các câu trả lời đúng, công thức (2) áp dụng cho các câu trả lời sai. Khi sinh viên bắt đầu kiểm tra, Giá_trị_cũ là giá trị đạt được của sinh viên

trước khi làm câu hỏi tiếp theo. Sau đó giá trị này sẽ được cập nhật theo phương án trả lời cho mỗi câu hỏi. Điểm số của bài kiểm tra trong khoảng từ 0 đến 1 [12].

Bài kiểm tra thích ứng còn được gọi là bài kiểm tra tuần tự là phương pháp đánh giá lựa chọn liên tiếp các câu hỏi từ ngân hàng câu hỏi phù hợp với năng lực của thí sinh trong suốt quá trình làm bài. Hai cách thức phổ biến nhất trong thiết kế bài kiểm tra tuần tự là kiểm tra thích ứng trên máy tính (computerized adaptive testing - CAT) và kiểm tra thích ứng đa giai đoạn (multistage adaptive testing - MST). Hai cách thức này được ứng dụng nhiều trong đo lường và đánh giá trong giáo dục, đo lường chất lượng cuộc sống trong y tế,... CAT ứng dụng trong các bài kiểm tra ACCUPLACER, GMAT và ASVAB, MST ứng dụng trong các bài kiểm tra GRE, AICPA và đo lường quá trình học tập. Trong cả hai cách thức CAT và MST, các câu hỏi được chọn liên tiếp dựa trên ước lượng năng lực hiện tại của thí sinh trong quá trình làm bài [13]. Ngoài ra, MST còn sử dụng để định cỡ câu hỏi [3].

CAT được phát triển bởi Lord (1971), Owen (1975), Weiss và các cộng sự (1976) để đáp ứng liên tục độ khó câu hỏi đối với năng lực của thí sinh "tốc hành" trong quá trình kiểm tra. Các thuật toán chọn câu hỏi và các phương pháp ước lượng năng lực thí sinh là 2 chủ đề nghiên cứu chính của CAT. Ngày nay, các thuật toán và phương pháp này liên tục được cải tiến sao cho đảm bảo chọn câu hỏi trong ngân hàng câu hỏi, ước lượng năng lực của thí sinh hiệu quả, phù hợp với năng lực của thí sinh và trong thời gian ngắn nhất [13]. CAT chỉ được triển khai khi ngân hàng câu hỏi đủ lớn và các câu hỏi được định cỡ [10].

*** Ưu điểm của CAT:**



Hình 2: Mô hình kiểm tra thích ứng 2 giai đoạn 1- 4 và 3 giai đoạn 1-2-4

Mỗi cá nhân thực hiện một bài kiểm tra phù hợp với năng lực nhất, nội dung và độ dài của bài kiểm tra có thể khác nhau giữa các thí sinh.

CAT được tối ưu hóa cho mỗi cá nhân như hiệu quả đo lường được nâng lên do có thể sử dụng số lượng câu hỏi ít hơn yêu cầu để đạt được độ chính xác tương tự và mỗi thí sinh có thể được thử thách ở cấp độ phù hợp với năng lực bản thân, khi đó thí sinh có thêm động lực và tích lũy kinh nghiệm.

Số lượng câu hỏi tối thiểu phù hợp nhất với năng lực mỗi thí sinh. Do đó, việc xây dựng bài kiểm tra, tổ chức kiểm tra có thể tiết kiệm thời gian và tiền bạc.

Do các câu hỏi và các bài kiểm tra được phát triển trong khuôn khổ lý thuyết ứng đáp câu hỏi, bài kiểm tra có một số đặc trưng biết trước và do đó chắc chắn chất lượng đo lường tốt hơn.

*** Lợi thế của CAT:**

- Phản hồi tích cực, cho điểm tích cực và khách quan.
- Các dạng câu hỏi mới có giá trị, phong phú và xác thực hơn.
- Quản lý bài kiểm tra hiệu quả hơn.
- Khả năng linh hoạt về thời gian và địa điểm kiểm tra.
- Bảo mật bài kiểm tra tốt hơn.
- Tạo động lực cho thí sinh hơn.
- Khả năng ứng dụng lý thuyết trắc nghiệm hiện đại cao [3].

Trong cả CAT và MST, các câu hỏi được lựa chọn liên tục dựa trên các ước lượng năng lực hiện tại của thí sinh. Tuy nhiên, MST khác với CAT về số câu hỏi được thực hiện tại một thời điểm. Trước kia, khi thí sinh hoàn thành một nhóm câu hỏi, nhóm tiếp theo được chọn dựa trên năng lực của thí sinh ở thời điểm hiện tại. Do đó, dạng câu hỏi đã lựa chọn phải được xác định độ khó. Sau này, chỉ một câu hỏi được chọn tại một thời điểm. Trong trường hợp này, CAT có thể được xem như bài kiểm tra thích ứng dựa trên độ khó câu hỏi, trong khi MST được xem như bài kiểm tra thích ứng dựa trên mô-đun. Trong MST, các câu hỏi được nhóm thành các cấp độ nhằm đến các mức năng lực khác nhau và đặt ngang hàng với các cấp độ khác nhau dựa vào các yêu cầu như nội dung, đặc trưng kiểm tra và các mục tiêu thống kê (Hình 2).

Ước lượng ban đầu của một thí sinh thu được dựa vào các câu trả lời ở giai đoạn đầu (giai đoạn định hướng/hướng), sau đó thí sinh được định hướng đến mô-đun có độ khó thích hợp nhất ở giai đoạn tiếp theo. Ước lượng năng lực cuối cùng của thí sinh dựa vào các câu trả lời của toàn bộ bài kiểm tra. Do đó, thành phần quan trọng của khung MST là cấu trúc các “cấp độ” khác nhau ngang hàng [13].

Các nghiên cứu hiện nay chủ yếu ứng dụng lý thuyết ứng đáp câu hỏi kết hợp với các thuật toán khác

nhau để xây dựng và triển khai ngân hàng câu hỏi nhằm tối ưu hóa bài kiểm tra đánh giá phù hợp với năng lực thí sinh nhất. Hơn nữa, phương pháp kiểm tra, đánh giá triển khai phải đảm bảo tính giáo dục, công bằng, khách quan, toàn diện, công khai và phát triển, đặc biệt phải đánh giá được các năng lực khác nhau của thí sinh.

5. Kết luận

Chất lượng ngân hàng câu hỏi là yếu tố chính yếu trong đánh giá năng lực thích ứng. Vì vậy, định cỡ các câu hỏi trong ngân hàng câu hỏi là quá trình tốn nhiều thời gian và công sức của các nhà chuyên môn, các chuyên gia máy tính và sự tham gia của thí sinh trong các kì thi. Quá trình định cỡ câu hỏi được thực hiện liên tục không chỉ đảm bảo chuẩn xác các tham số câu hỏi mà còn đáp ứng sự thay đổi và phát triển năng lực của thí sinh, sự phát triển của xã hội. Từ một số phương pháp, giải pháp xây dựng và triển khai ngân hàng câu hỏi vào các kì thi đánh giá năng lực, các trường có thể nghiên cứu, kết hợp áp dụng vào thực tiễn trong kiểm tra đánh giá năng lực người học trong xu hướng đổi mới đánh giá hoạt động học tập theo cách tiếp cận năng lực và triết lý đánh giá vì sự tiến bộ trong học tập.

TÀI LIỆU THAM KHẢO

[1]. Sands W., Waters B., McBride, (1997), *Computerized Adaptive Testing: From Inquiry to Operation*, American Psychological Association, Washington, DC, USA.

[2]. Prometric: <https://www.prometric.com/en-us/our-solutions/test-development/pages/item-banking.aspx>.

[3]. Educational Testing Service, (2006), *Hawaii Formative Assessment Item Bank*, ETS.

[4]. Nguyễn Công Khanh, Đào Thị Oanh, Lê Mỹ Dung, (2014), *Kiểm tra đánh giá trong giáo dục*, NXB Đại học Sư phạm, Hà Nội.

[5]. Oxford University: <http://global.oup.com/uk/orc/learntestbanks/>.

[6]. Lâm Quang Thiệp, (2008), *Trắc nghiệm và ứng dụng*, NXB Khoa học Kỹ thuật, Hà Nội.

[7]. Rudner L., (1998), *Practical Assessment, Research & Evaluation*, 6(4). Available online: <http://PAREonline.net/getvn.asp?v=6&n=4>.

[8]. Burghof K. L., (2001), *Assembling an item-bank for computerised linear and adaptive testing in Geography*, International Education Journal, Vol 2, No 4, pp. 74 - 83.

[9]. López-Cuadrado J., Armendariz A., Pérez T. A., Arruabarrena R., (2008), *Helping tools for item bank calibration and development of computerized adaptive tests*, *Proceeding of International Technology, Education and Development Conference (INTED'08)*.

[10]. Molina J. G., Pareja I., Sanmartín J., (2008), *Modeling item banking: Analysis and design of a*



computerized system, Revista Electrónica de Metodología Aplicada, 13(2), pp. 1-14.

[11]. Aminifar E., Alipour M. ,(2014), Developing an item bank for homogeneous second order differential equations by calibrated items, European Journal of Educational Sciences, 1(2), pp. 161 - 170.

[12]. Fong A. T., Siew H. H., Yee P. L., Sun L. C., (2006), *Intelligent Question Bank and Examination System*, Proceedings of the 5th WSEAS International Conference on E-ACTIVITIES, Venice, Italy, November 20-22, pp. 142-147.

[13]. Wang S., Lin H., Chang H., Douglas J. , (2016), *Hybrid Computerized Adaptive Testing: From Group Sequential Design to Fully Sequential Design*, Journal of Educational Measurement, 53(1).

[14]. Lâm Quang Thiệp, (2011), *Đo lường trong giáo dục: Lý thuyết và ứng dụng*, NXB Đại học Quốc gia Hà Nội.

[15]. Eggen T. J. H. M., (2007), *Choices in CAT Models in the Context of Educational Testing*, Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing.

STUDY ON QUESTION BANK TO ASSESS CANDIDATE COMPETENCIES

Do Dinh Thai - Email: dodinhthai@sgu.edu.vn

Le Chi Lan - Email: chilanih@yahoo.com

Saigon University

Abstract: *The question banks play an important role in evaluating the training quality in terms of the trainee's knowledge, skills and attitudes as well as in assessing the candidate's competencies in an exam. To assess correctly the candidate's competencies, quality question bank is an indispensable component. Based on the analysis of the interests of the question bank, the challenges to the question bank, the article examines some methods of question sizing, ways of developing and implementing methods to assess the adaptation abilities through a question bank in which the items have been sized to assess and categorize candidates' competencies in examinations. In particular, the method of question sizing determines the quality of question banks while the method of assessing adaptive ability is to optimize the time and organization of the assessment, but still ensure the classification of candidates' competencies. Through the above approach, universities can apply in the assessment of candidates' competencies in examinations.*

Keywords: *Question bank; assess; competence; student.*