

Phát triển ngân hàng trắc nghiệm thích ứng để đánh giá năng lực đọc hiểu môn Ngữ văn của học sinh lớp 10 trung học phổ thông

Lê Thái Hưng¹, Trần Thị Hoa²,
Đặng Thị Mây³, Hoàng Lan Hương⁴

¹ Email: hunglethai82@gmail.com

² Email: tranhoak56@gmail.com

³ Email: dangmaykhgd@gmail.com

⁴ Email: huonghoangsp@gmail.com

Trường Đại học Giáo dục - Đại học Quốc gia Hà Nội
144 Xuân Thủy, Cầu Giấy, Hà Nội, Việt Nam

TÓM TẮT: Trắc nghiệm thích ứng trên máy tính (Computerized Adaptive Testing - CAT) là một hình thức kiểm tra đánh giá cho phép rút ngắn số lượng câu hỏi nhưng vẫn đảm bảo độ chính xác về đánh giá năng lực của thí sinh. Một trong những phần cốt lõi của hệ thống trắc nghiệm thích nghi là các thuật toán ước lượng năng lực thí sinh và lựa chọn câu hỏi. Các thuật toán này đóng vai trò đầu máy trong quá trình vận hành hệ thống trắc nghiệm thích nghi trên máy tính. Nghiên cứu này sẽ phát triển các thuật toán cốt lõi trong hệ thống trắc nghiệm thích nghi, từ đó lập trình hệ thống trắc nghiệm thích ứng. Nghiên cứu cũng tiến hành xây dựng ngân hàng gồm 500 câu hỏi trắc nghiệm thích ứng được chuẩn hoá theo lý thuyết IRT với điều kiện độ khó tuân theo phân phối chuẩn thoả mãn kiểm định Kolmogorov-Smirnov, để đánh giá năng lực đọc hiểu của học sinh lớp 10. Kết quả vận hành thử nghiệm với hệ thống ngân hàng câu hỏi bước đầu cho thấy: Bộ câu hỏi xây dựng đã đáp ứng yêu cầu mô hình ước lượng năng lực và thuật toán cốt lõi đáp ứng được yêu cầu của trắc nghiệm thích ứng.

TỪ KHÓA: Trắc nghiệm thích ứng trên máy tính; đánh giá năng lực; năng lực đọc hiểu.

→ Nhận bài 25/10/2019 → Nhận kết quả phản biện và chỉnh sửa 05/12/2019 → Duyệt đăng 25/12/2019.

1. Đặt vấn đề

Trắc nghiệm thích ứng là thuật ngữ chỉ một phương pháp đánh giá (ĐG) thí sinh trong đó có học sinh (HS), sinh viên, bệnh nhân... bằng hình thức kiểm tra trắc nghiệm nhưng ĐG theo hướng năng lực (NL) của thí sinh bằng bộ câu hỏi tương ứng với mức NL đó. Hệ thống trắc nghiệm thích ứng là một hệ thống phần mềm được phát triển trên cơ sở mô hình trắc nghiệm thích ứng để ĐG thí sinh. Lần đầu tiên, hệ thống cung cấp cho thí sinh một câu hỏi vừa đủ khó đối với thí sinh. Trắc nghiệm thích ứng giúp cho việc ĐG phù hợp với NL của người học theo từng giai đoạn. Trong nghiên cứu này, chúng tôi lựa chọn vấn đề phát triển ngân hàng trắc nghiệm thích ứng để ĐG NL đọc hiểu của HS lớp 10 THPT. *Kết quả nghiên cứu trình bày ở đây được tài trợ bởi nhóm đề tài nghiên cứu khoa học có mã số QS.17.14 và QS.17.15 của Trường Đại học Giáo dục được thực hiện bởi nhóm nghiên cứu của Khoa Quản trị Chất lượng, Trường Đại học Giáo dục.*

2. Nội dung nghiên cứu

2.1. Năng lực đọc hiểu và thang đánh giá

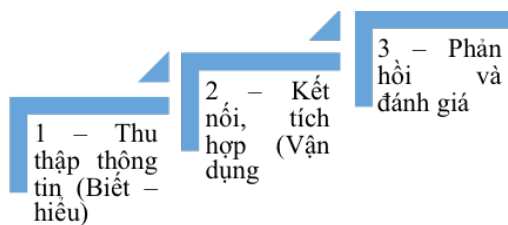
Do chi phối bởi mục tiêu môn học và điều kiện thực hiện, ĐG kết quả học tập của HS trong môn Ngữ văn hiện nay tập trung chủ yếu vào hai NL đọc và viết. Trong bài báo này, chúng tôi xây dựng bộ công cụ ĐG NL đọc hiểu. Đọc hiểu là một NL tiếp nhận văn bản, thông qua hoạt động của con người đọc chữ, xem các kí hiệu bằng biểu, hình ảnh trong

nhều loại văn bản khác nhau, nhằm xử lí thông tin trong văn bản để phục vụ những mục đích cụ thể trong học tập hoặc giải quyết các nhiệm vụ thực tiễn của cuộc sống. Đọc hiểu luôn là một nội dung trọng tâm trong chương trình môn Ngữ văn ở trường phổ thông của hầu hết các quốc gia trên thế giới trong thời kì từ đầu thế kỉ XX đến nay.

Trong dạy học môn Ngữ văn, NL đọc hiểu văn bản văn học rất được coi trọng. Việc ĐG NL đọc hiểu của HS hiện nay thường diễn ra dưới hai hình thức: Kiểm tra miệng (Yêu cầu HS nhắc lại một nội dung nào đó của bài học đã ghi chép trong vở) và kiểm tra viết (Viết về một vấn đề thuộc phương diện nội dung hoặc nghệ thuật của văn bản đã học). Những nhiệm vụ này chưa ĐG được NL đọc hiểu các loại văn bản khác nhau của HS. Vì vậy, vấn đề đặt ra là cần đổi mới ĐG NL đọc hiểu của HS bằng việc sử dụng những văn bản mới (Bao gồm cả văn bản văn học và văn bản thông tin, có cùng đề tài, chủ đề hoặc thể loại với văn bản đã học trong chương trình, sách giáo khoa), yêu cầu HS vận dụng những kiến thức, kĩ năng đã có vào việc đọc hiểu và cảm thụ văn bản mới này. Các câu hỏi ĐG NL đọc hiểu nên được thiết kế theo cách làm của Chương trình ĐG HS quốc tế (viết tắt là PISA), bao gồm: câu hỏi mở, câu hỏi đóng, câu hỏi trắc nghiệm khách quan nhiều lựa chọn, câu hỏi có - không, đúng - sai phức hợp. ĐG NL đọc hiểu của HS phải được tiến hành thường xuyên trong các bài kiểm tra từ 1-2 tiết, bài kiểm tra học kì, kiểm tra cuối năm, kì thi THPT cấp quốc gia. Trên cơ sở thang NL 6 mức của PISA có thể nhận

ra mỗi liên hệ với thang 4 mức mà hiện nay chúng ta đang sử dụng. Các mức độ này có thể được mô tả như sau gồm: Nhận biết, thông hiểu, vận dụng, vận dụng cao. Đọc hiểu mỗi kiểu văn bản và thể loại nói chung có các yêu cầu cần đạt sau (xem Bảng 1):

Trong thực tế, việc xếp một hành vi nhận thức vào một trong bốn cấp độ nhận thức trên không hề đơn giản. Các nhà khoa học đều nhận thấy có sự giao thoa nhất định giữa các bậc nhận thức liên nhau. Trong các cấp độ nhận thức, cấp độ phân tích, ĐG, sáng tạo vận dụng được các nhà khoa học xếp vào nhóm các cấp độ tư duy bậc cao. Để thuận tiện cho việc sử dụng thang sáu bậc nhận thức của Bloom để xây dựng mục tiêu dạy học và ĐG việc thực hiện mục tiêu, có một cách nhóm các cấp độ nhận thức thành 3 bậc đã và đang được sử dụng phổ biến (xem Hình 1).



Hình 1: Cấp độ NL đọc hiểu (OECD) [1, tr.60]

Từ những vấn đề lí thuyết liên quan đến NL đọc hiểu, cấu trúc của NL đọc hiểu, các thành tố, kĩ năng của NL đọc hiểu, chuẩn nội dung của Chương trình Ngữ văn lớp 10 THPT chúng tôi lựa chọn xây dựng khung ĐG NL đọc hiểu theo 3 cấp độ như trên. Trong nghiên cứu này, nhóm tác giả chúng tôi phát triển ngân hàng câu hỏi trắc nghiệm theo các cấp độ hướng đến các thành tố của NL đọc hiểu. Nghiên cứu xây dựng câu hỏi ở cấp độ 1 - Thu thập thông tin, HS đọc hiểu những kiến thức về hình thức (từ dễ đến khó); Cấp độ 2 - Kết nối, tích hợp ngân hàng trắc nghiệm hướng đến việc xây dựng những câu hỏi đọc hiểu về nội dung; Cấp độ 3 - Phản hồi và ĐG những câu hỏi ở cấp độ này chủ yếu kiểm tra về khả năng đọc mở rộng, liên kết, so sánh. Ở từng cấp độ, chúng tôi lại chia theo ba mức dễ, trung bình, khó để ĐG cụ thể, chi tiết NL đọc hiểu của HS thông qua đề khảo sát trắc nghiệm thích ứng.

2.2. Giới thiệu về trắc nghiệm thích ứng

Để ĐG NL đọc hiểu của HS phù hợp với từng mức độ khác nhau, việc xây dựng ngân hàng câu hỏi trắc nghiệm thích ứng là cần thiết. Trên thế giới, đã có nhiều nghiên

Bảng 1: Các thành tố của NL đọc hiểu

Yêu cầu cần đạt	Nội dung cần đạt
Đọc hiểu hình thức	Thể hiện qua đặc điểm các kiểu văn bản và thể loại, các thành tố của mỗi kiểu văn bản và thể loại (câu chuyện, cốt truyện, truyện kể, nhân vật, không gian, thời gian, người kể chuyện, điểm nhìn, văn thơ, nhật thơ, lí lẽ, bằng chứng...), ngôn ngữ biểu đạt,...
Đọc hiểu nội dung	Thể hiện qua chi tiết, đề tài, chủ đề, tư tưởng, thông điệp.
Liên hệ, so sánh	Kết nối giữa các văn bản, văn bản với bối cảnh lịch sử, văn hoá, xã hội, kết nối văn bản với trải nghiệm cá nhân người đọc; đọc hiểu văn bản đa phương thức,...

cứ cho thấy hiệu quả trong việc sử dụng CAT trong ĐG người học. Một số kì thi sử dụng CAT: Trắc nghiệm thích ứng toán học (CAT-Math) và đọc (CAT-Reading) ... Một trong những ưu thế của CAT là chúng ta có thể tổ chức ĐG đồng thời trên diện rộng với số lượng lớn HS tham gia làm bài trên hệ thống máy tính được kết nối mạng. Hơn nữa, CAT cho phép phân tích các chỉ số về NL của thí sinh ngay sau khi thí sinh trả lời câu hỏi và thông tin về NL của HS được cập nhật thường xuyên trong quá trình làm bài cho đến khi đo được NL thực sự của họ. Vì vậy, CAT không những giúp ĐG chính xác NL mà còn đưa ra thông tin đầy đủ và toàn diện về NL HS cho cơ sở giáo dục triển khai các mô hình học tập thích ứng. Do vậy, trong mô hình Trắc nghiệm thích ứng: Thuật toán lựa chọn câu hỏi tiếp theo phù hợp với khả năng hiện tại của thí sinh là khó khăn vì phải được tính toán một cách tối ưu nhất. Bên cạnh đó, một ngân hàng câu hỏi được chuẩn hoá theo lí thuyết ứng đáp câu hỏi cần được xây dựng. Số câu hỏi trong ngân hàng cần đủ lớn để đạt được phân bố chuẩn với tham số độ khó.

2.3. Xây dựng bảng đặc tả và viết câu hỏi

2.3.1. Bảng đặc tả đề thi đánh giá năng lực đọc hiểu (xem Bảng 2)

2.3.2. Viết câu hỏi và thử nghiệm

Với những thuận lợi của hình thức ĐG bằng trắc nghiệm thích ứng với môn Ngữ văn giúp HS nhận ra NL của bản thân đang ở mức độ nào, giáo viên sẽ có những hình thức và phương pháp dạy học phù hợp. Khi HS được trang bị kiến thức đầy đủ sẽ tiếp tục chinh phục thang NL ĐG ở mức độ cao nhất, chúng tôi tiến hành xây dựng ngân hàng câu hỏi theo bảng đặc tả và ba cấp độ của NL đọc hiểu. Phân tích một số câu hỏi minh họa theo ba cấp độ (lựa chọn những câu hỏi ngoài chương trình để phân tích, làm rõ NL của HS).

Mức 1: Thu thập thông tin

Với những câu hỏi ở mức 1 sẽ kiểm tra các kiến thức liên quan đến NL đọc hiểu hình thức (phương thức biểu đạt, các biện pháp tu từ...)

Độc đoạn thơ sau và trả lời câu hỏi từ câu 43 đến câu 44:

Bao giờ cho tới mùa thu
 Trái hồng trái bưởi đánh đu giữa rằm
 Bao giờ cho tới tháng năm
 Mẹ ra trái chiếu ta nằm đếm sao.
 Ngân Hà chảy ngược lên cao
 Quạt mo vỗ khúc nghêu ngoao thẳng Bờm...
 Bờ ao đom đóm chập chờn
 Trong leo lẻo những vui buồn xa xôi.

Bảng 2: Bảng đặc tả đề thi ĐG NL đọc hiểu (Trọng số điểm từng câu như nhau)

Nội dung Kiến thức	Tiểu nội dung	Thu thập thông tin	Kết nối, tích hợp	Phản hồi ĐG
Phương thức biểu đạt	Tự sự, miêu tả, biểu cảm, thuyết minh, nghị luận	Phương thức biểu đạt (văn bản ngoài sách giáo khoa).		
Tiếng Việt	Biện pháp tu từ tích hợp kiến thức Tiếng Việt khác	Xác định đúng biện pháp tu từ; Phân tích hiệu quả của biện pháp tu từ.		
Lịch sử văn học	Văn bản thông tin	Mô tả được các bộ phận hợp thành của văn học Việt Nam, hệ thống thể loại, 2 thành phần chủ yếu, 4 giai đoạn phát triển, 3 đặc điểm lớn về nghệ thuật, 3 đặc điểm lớn về nội dung của văn học trung đại Việt Nam.	Phân tích con người Việt Nam; Phân biệt tự sự dân gian - trữ tình dân gian - sân khấu dân gian; đặc điểm cơ bản về nội dung và nghệ thuật của từng giai đoạn; biểu hiện cơ bản của chủ nghĩa yêu nước, chủ nghĩa nhân đạo và cảm hứng thế sự trong văn học trung đại Việt Nam.	
Tự sự dân gian	Sử thi, truyền thuyết, truyện cổ tích, truyện cười	Khái niệm, đặc điểm thể loại truyền thuyết, cổ tích, truyện cười; Chọn sự việc, chi tiết tiêu biểu.	Phân tích được nội dung và nghệ thuật, nhân vật trong truyền thuyết, truyện cổ tích, truyện cười.	
Trữ tình dân gian	Ca dao than thân, yêu thương, tình nghĩa, ca dao hài hước.	Nêu được khái niệm, đặc điểm cơ bản về nội dung và hình thức của ca dao.	Phân tích được các bài ca dao than thân, yêu thương, tình nghĩa, ca dao hài hước.	
Thơ	Các tác phẩm thơ Việt Nam và nước ngoài	Xác định thể thơ, những nét đặc trưng về tác giả, tác phẩm.	Hình tượng con người và quân đội thời Trần, chí làm trai, bức tranh thiên nhiên, tâm trạng con người.	
Thể phú	Phú sông Bạch Đằng	Nêu được đặc điểm thể phú; tiểu sử, sự nghiệp văn học của tác giả.		
Thể cáo	Đại cáo bình Ngô	Nêu được đặc điểm thể cáo; tiểu sử, sự nghiệp văn học của tác giả.		Tư tưởng nhân nghĩa của Nguyễn Trãi.
Tự sự trung đại	Chuyện chức phán sự đền Tản Viên; Tam quốc diễn nghĩa	Nêu được đặc điểm thể truyện kì; tiểu thuyết cổ điển Trung Hoa, tiểu sử, sự nghiệp văn học của tác giả.	Phân tích nhân vật, tình huống truyện.	Bình giá được nghệ thuật xây dựng nhân vật.
Thể ngâm khúc	Tình cảnh lẻ loi của người chinh phụ		Phân tích tâm trạng của người chinh phụ.	
Truyện thơ	Các văn bản trong Truyện Kiều	Nội dung và hình thức của truyện thơ bác học, những nét chính về tiểu sử, sự nghiệp sáng tác văn học của Nguyễn Du.		So sánh Truyện Kiều với Kim Vân Kiều truyện, "Kính gửi cụ Nguyễn Du", lí giải mối quan hệ giữa thời đại và hình tượng nhân vật.
Đọc hiểu liên văn bản	Các loại văn bản ngoài chương trình	Nhận biết, gọi tên kiến thức về thể loại, kiểu văn bản, biện pháp nghệ thuật.	Những kiến thức liên quan đến kết nối thông tin, phân tích, cảm nhận nội dung.	ĐG, so sánh vấn đề thông qua những câu hỏi hình ảnh, tư liệu mở rộng.

Mẹ ru cái lẽ ở đời
 Sữa nuôi phần xác, hát nuôi phần hồn
 Bà ru mẹ... Mẹ ru con
 Liệu mai sau các con còn nhớ chăng.

(Trích Ngồi buồn nhớ mẹ ta xưa, Nguyễn Duy,
 Mẹ và em, NXB Thanh Hóa, 1987)

K2. Câu 44. Hai biện pháp tu từ được sử dụng trong bốn câu đầu của đoạn thơ trên là:

- A. Điệp cấu trúc, nhân hóa
- B. Liệt kê, ẩn dụ
- C. Điệp cấu trúc, ẩn dụ
- D. Ẩn dụ, nói quá

K2 là câu hỏi có độ khó trung bình ở mức 1, HS dựa vào những dấu hiệu nhận biết biện pháp tu từ để xác định đáp án chính xác.

Mức 2: Kết nối, tích hợp

Đọc bài thơ sau và trả lời câu hỏi từ câu 49 đến câu 51:

Tôi yêu truyện cổ nước tôi
 Vừa nhân hậu lại tuyệt vời sâu xa
 Thương người rồi mới thương ta
 Yêu nhau dù mấy cách xa cũng tìm
 Ở hiền thì lại gặp hiền
 Người ngay thì gặp người tiên độ trì
 Mang theo truyện cổ tôi đi

Như trong cuộc sống thăm thì tiếng xưa
 Vàng con nắng, trắng con mưa
 Con sông chảy có rặng dừa nghiêng soi
 Đòi cha ông với đòi tôi
 Như con sông với chân trời đã xa
 Chỉ còn truyện cổ thiết tha
 Cho tôi nhận mặt ông cha của mình
 Rất công bằng, rất thông minh
 Vừa độ lượng lại đa tình, đa mang.
 Thị thơm thì giấu người thơm
 Chăm làm thì được áo cơm cửa nhà
 Đẽo cày theo ý người ta
 Sẽ thành khúc gỗ chẳng ra việc gì
 Tôi nghe chuyện cổ thăm thì
 Lời cha ông dạy cũng vì đòi sau
 Đậm đà cái tích trâu cau
 Miếng trâu đỏ thắm nặng sâu tình người
 Sẽ đi qua cuộc đời tôi
 Bấy nhiêu thời nữa chuyện đời xa xôi
 Nhưng bao chuyện cổ trên đời
 Vẫn luôn mới mẻ rạng ngời lương tâm.

(Truyện cổ nước mình, Lâm Thị Mĩ Dạ)

A1. Câu 1. Bài thơ gợi nhắc đến những tác phẩm nào trong kho tàng văn học dân gian Việt Nam?

- A. Sự tích Trâu Cau, Sơn Tinh Thủy Tinh, Tấm Cám
- B. Đẽo cày giữa đường, Tấm Cám, Sự tích Bánh chưng Bánh dày
- C. Đẽo cày giữa đường, Tấm Cám, Sự tích Trâu Cau
- D. Tấm Cám, Thạch Sanh, Sự tích Trâu Cau

Câu hỏi A1 là câu hỏi dễ ở mức 2 sinh cần vận dụng những kiến thức đã học về văn học dân gian (những câu chuyện cổ tích, truyền thuyết đã học từ lớp 6 để tìm ra tên các tác phẩm).

Mức 3: Phản hồi và ĐG

R3. Câu 60. Dưới đây là bức ảnh Người đàn ông đang rơi (The falling man) được nhiếp ảnh gia Richard Drew chụp ngày 11 tháng 09 năm 2001 và đoạn trích trong bài thơ Tấm ảnh chụp ngày 11 tháng 9 của nhà thơ Wislawa Szymborska:

Họ vẫn còn trên không trung,
 trong những vị trí
 mở toang.
 Chỉ hai việc tôi có thể làm cho họ:
 miêu tả chuyến bay
 và không đặt tay
 ghi vào câu kết.

(Trích Các nhà thơ giải Nobel,
 NXB Lao động, 2007, Lê Bá Thự dịch)

Khi “miêu tả chuyến bay/ và không đặt tay/ ghi vào câu kết”, nhà thơ Wislawa Szymborska đã chia sẻ cảm xúc nào với người nghệ sĩ nhiếp ảnh trong khoảnh khắc bấm máy?

- A. Cảm nhận sự bất lực của bản thân và của nghệ thuật trước sự mong manh, phù du và nỗi đau của thân phận con người.
- B. Niềm kính hoàng trước cái chết của những người xung quanh
- C. Niềm hoang mang trước tương lai nhân loại phải đối



diện với chủ nghĩa khủng bố

D. Niềm trân trọng về sự mệnh của người nghệ sĩ trong việc phản ánh chân thực, kịp thời những đổi thay của đời sống xã hội.

Câu hỏi mức R3 là câu hỏi có độ khó cao nhất vì yêu cầu HS tổng hợp thông tin từ nhiều nguồn, so sánh, đối chiếu và lập luận để tìm ra đáp án chính xác nhất.

Dữ liệu thử nghiệm đề văn (đề gồm 60 câu hỏi trắc nghiệm khách quan) lần 1 với 157 HS tham gia, không có dự liệu missing. Kết quả cho thấy có 9 câu (▲) cần được điều chỉnh hoặc loại bỏ. Xem xét chi tiết từng câu, nhận thấy tất cả các câu hỏi đều không thỏa mãn độ phân biệt theo lý thuyết cổ điển ($Discr > 0.2$). Về độ khó, hầu hết các câu hỏi này cũng thuộc nhóm quá dễ (C1, C5, C8, C9, C11, C13, C16, $b < -3$). Biểu diễn phân tích nhân tố của các câu hỏi này cũng không đáp ứng yêu cầu, có sự chênh lệch lớn giữa đường lý thuyết và thực nghiệm. Những câu này cần loại bỏ hoặc điều chỉnh nhưng trong nghiên cứu này, chúng tôi loại bỏ để viết bằng câu hỏi mới và tiếp tục thử nghiệm cho đến khi ngân hàng câu hỏi đủ lớn và đáp ứng yêu cầu phân phối chuẩn về độ khó. Những điểm cần lưu ý khi phân tích cụ thể các câu hỏi không thỏa mãn được tổng hợp cho giai đoạn tiếp theo nhằm giảm thiểu số lượng câu hỏi không đạt yêu cầu. Các câu hỏi còn lại phù hợp với mô hình là những câu hỏi thường có độ khó nằm trong khoảng $[-3, +3]$, thỏa mãn điều kiện độ phân biệt theo lý thuyết cổ điển ($Discr > 0.2$) và chất lượng đáp án nhiều tốt (xem Bảng 3).

Để có nhận định chung về đề thi, chúng tôi tiến hành phân tích phổ điểm theo thang NL và hàm thông tin của đề thi (xem Biểu đồ 1). Kết quả cho thấy, NL trung bình của thí sinh là 0,07 gần với mức NL trung bình lý thuyết, phân bố NL của thí sinh có dáng điệu chuẩn và số câu hỏi đáp ứng mô hình là 55/60 câu. Hàm thông tin có đỉnh đạt xấp xỉ 13, độ tin cậy bằng 0.81 đáp ứng yêu cầu.

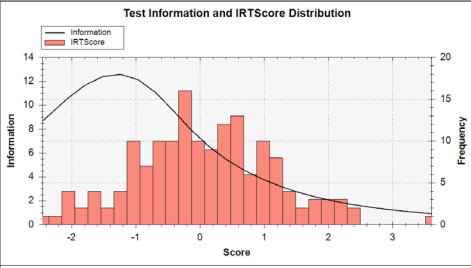
2.3.3. Xây dựng ngân hàng câu hỏi và kết quả thử nghiệm

Chúng tôi tiến hành phát triển ngân hàng câu hỏi (gồm 500 câu) đáp ứng các yêu cầu của trắc nghiệm thích ứng đề ĐG được NL HS ở mức phù hợp nhất. Tương tự với 11 đề thi môn Ngữ văn, ngân hàng 500 câu hỏi tiến hành thử nghiệm và thu được tham số có độ khó và độ phân biệt theo lý thuyết khảo thí IRT. Tiến hành kiểm nghiệm phân phối chuẩn với kiểm định Kolmogorov-Smirnov thu nhận được kết quả như sau (xem Biểu đồ 2):

Bảng 3: Kết quả phân tích sự phù hợp với lý thuyết IRT (items fit) và các tham số (tóm lược)

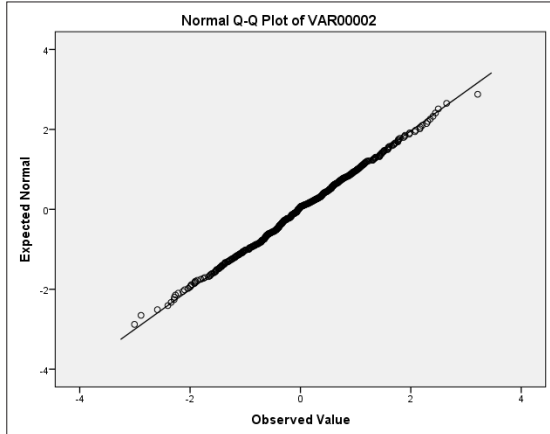
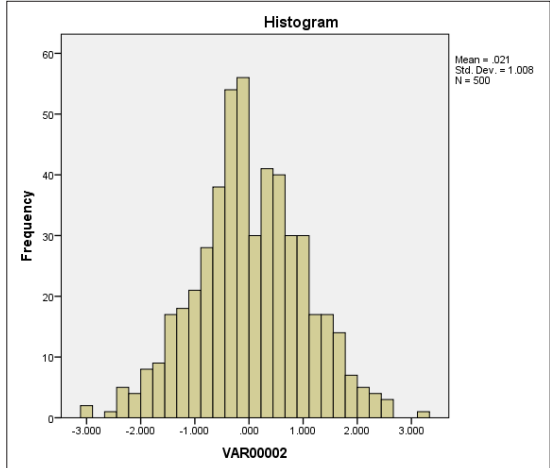
Use	O	Name	Discr	PVal	PBis	a	b	c
<input checked="" type="checkbox"/>	△	C5	0.05	0.35	0.01	-0.01	-53.48	0.00
<input checked="" type="checkbox"/>	△	C6	0.07	0.04	0.11	0.29	6.54	0.00
<input checked="" type="checkbox"/>	△	C8	0.10	0.87	0.14	0.18	-6.33	0.00
<input checked="" type="checkbox"/>	△	C10	0.10	0.93	0.24	0.42	-3.88	0.00
<input checked="" type="checkbox"/>	△	C45	0.10	0.78	0.16	0.20	-3.82	0.00
<input checked="" type="checkbox"/>	△	C24	0.11	0.42	0.13	0.14	1.37	0.00
<input checked="" type="checkbox"/>	△	C7	0.12	0.94	0.25	0.49	-3.74	0.00
<input checked="" type="checkbox"/>	△	C14	0.12	0.95	0.31	0.68	-3.01	0.00
<input checked="" type="checkbox"/>	△	C16	0.12	0.89	0.13	0.14	-8.55	0.00
<input checked="" type="checkbox"/>	△	C11	0.15	0.65	0.14	0.08	-4.44	0.00
<input checked="" type="checkbox"/>	△	C9	0.16	0.60	0.12	0.03	-7.52	0.00
<input checked="" type="checkbox"/>	△	C32	0.16	0.94	0.36	0.89	-2.46	0.00
<input checked="" type="checkbox"/>	△	C1	0.19	0.82	0.14	0.09	-9.57	0.00
<input checked="" type="checkbox"/>	△	C23	0.19	0.83	0.25	0.28	-3.45	0.00
<input checked="" type="checkbox"/>	△	C36	0.19	0.27	0.20	0.20	2.97	0.00
<input checked="" type="checkbox"/>	△	C13	0.20	0.69	0.18	0.14	-3.31	0.00
<input checked="" type="checkbox"/>	△	C25	0.21	0.84	0.35	0.59	-1.96	0.00
<input checked="" type="checkbox"/>	△	C34	0.21	0.92	0.36	0.81	-2.33	0.00
<input checked="" type="checkbox"/>	△	C42	0.22	0.24	0.19	0.29	2.54	0.00
<input checked="" type="checkbox"/>	△	C2	0.23	0.34	0.24	0.27	1.52	0.00
<input checked="" type="checkbox"/>	△	C47	0.23	0.88	0.35	0.85	-1.84	0.00
<input checked="" type="checkbox"/>	△	C12	0.26	0.90	0.37	0.71	-2.24	0.00
<input checked="" type="checkbox"/>	△	C18	0.26	0.85	0.33	0.48	-2.35	0.00
<input checked="" type="checkbox"/>	△	C20	0.26	0.76	0.31	0.37	-1.99	0.00
<input checked="" type="checkbox"/>	△	C22	0.26	0.90	0.37	0.66	-2.34	0.00

Use	O	Name	Discr	PVal	PBis	a	b	c
<input checked="" type="checkbox"/>	△	C43	0.31	0.80	0.37	0.69	-1.45	0.00
<input checked="" type="checkbox"/>	△	C59	0.31	0.62	0.25	0.30	-0.99	0.00
<input checked="" type="checkbox"/>	△	C21	0.32	0.50	0.27	0.24	0.03	0.00
<input checked="" type="checkbox"/>	△	C29	0.33	0.83	0.40	0.73	-1.65	0.00
<input checked="" type="checkbox"/>	△	C44	0.33	0.69	0.31	0.53	-1.02	0.00
<input checked="" type="checkbox"/>	△	C49	0.33	0.30	0.30	0.34	1.59	0.00
<input checked="" type="checkbox"/>	△	C53	0.33	0.85	0.47	1.21	-1.38	0.00
<input checked="" type="checkbox"/>	△	C3	0.35	0.80	0.35	0.57	-1.70	0.00
<input checked="" type="checkbox"/>	△	C58	0.35	0.27	0.30	0.39	1.61	0.00
<input checked="" type="checkbox"/>	△	C40	0.36	0.64	0.36	0.44	-0.85	0.00
<input checked="" type="checkbox"/>	△	C17	0.38	0.80	0.36	0.50	-1.87	0.00
<input checked="" type="checkbox"/>	△	C31	0.40	0.69	0.37	0.54	-1.05	0.00
<input checked="" type="checkbox"/>	△	C56	0.40	0.73	0.32	0.50	-1.32	0.00
<input checked="" type="checkbox"/>	△	C27	0.41	0.39	0.26	0.28	1.00	0.00
<input checked="" type="checkbox"/>	△	C4	0.43	0.46	0.32	0.37	0.29	0.00
<input checked="" type="checkbox"/>	△	C26	0.43	0.62	0.37	0.45	-0.71	0.00
<input checked="" type="checkbox"/>	△	C50	0.44	0.75	0.42	0.84	-1.05	0.00
<input checked="" type="checkbox"/>	△	C54	0.44	0.82	0.46	1.01	-1.27	0.00
<input checked="" type="checkbox"/>	△	C60	0.45	0.60	0.36	0.54	-0.51	0.00
<input checked="" type="checkbox"/>	△	C35	0.50	0.48	0.37	0.59	0.08	0.00
<input checked="" type="checkbox"/>	△	C38	0.52	0.50	0.40	0.67	0.01	0.00
<input checked="" type="checkbox"/>	△	C46	0.52	0.64	0.39	0.60	-0.67	0.00
<input checked="" type="checkbox"/>	△	C55	0.58	0.77	0.58	1.43	-0.90	0.00
<input checked="" type="checkbox"/>	△	C19	0.64	0.41	0.47	0.91	0.32	0.00
<input checked="" type="checkbox"/>	△	C52	0.65	0.73	0.59	1.35	-0.74	0.00



Mean	.07
StDev	.08
Skewness	.17
Kurtosis	.11
InterQuartileRange	.40
P-25	0.66
Median	0.01
P-75	.74
ResponseRate	.00
Reliability	.81
#Respondents	57.00
#Items	0.00
#OkayItems	5.00

Biểu đồ 1: Phân bố điểm NL và hàm thông tin của đề thi theo lý thuyết IRT



Descriptives

		Statistic	Std. Error
VAR00002	Mean	.02145	.045069
	95% Confidence Interval for Mean	Lower Bound: .06710 Upper Bound: .11000	
	5% Trimmed Mean	.02291	
	Median	.03700	
	Variance	1.016	
	Std. Deviation	1.007779	
	Minimum	3.008	
	Maximum	3.212	
	Range	6.220	
	Interquartile Range	1.300	

Tests of Normality

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
VAR00002	.037	500	.140	.998	500	.934

Biểu đồ 2: Kết quả chạy kiểm định Kolmogorov-Smirnov giá trị độ khó của 500 câu hỏi

Nhìn vào Biểu đồ 2, ta thấy sự phân bố độ khó và biểu đồ xác suất chuẩn Q-Q (Normal QQ plot) đều thể hiện đặc trưng của phân phối chuẩn (phân bố hình chuông, các trị số quan sát và mong đợi đều nằm trên đường chéo). Kết quả kiểm định phân phối chuẩn (Test of Normality) trường hợp dữ liệu lớn hơn 30 qua hệ số kiểm định Kolmogorov-Smirnov cho thấy hệ số ý nghĩa (Sig) = 0,119 > 0,05 nên giả thuyết H_0 thỏa mãn, độ khó của 500 câu hỏi tuân theo phân phối chuẩn. Với một ngân hàng câu hỏi được xây dựng đạt chuẩn chất lượng, hệ thống trắc nghiệm thích ứng cho phép rút ngắn thời gian làm bài của thí sinh mà vẫn đạt được độ chuẩn xác. Hệ thống trắc nghiệm này còn có các tính năng giúp giáo viên có thể dễ dàng quản lý ngân hàng câu hỏi, quản lý thí sinh và thu thập kết quả thi của TS một cách tự động.

3. Kết luận

Sau khi thử nghiệm ngân hàng 500 câu hỏi trắc nghiệm DG NL đọc hiểu môn Ngữ văn, so sánh với kết quả mô phỏng cho thấy kết quả chạy trên hệ thống trắc nghiệm thích ứng hoàn toàn trùng khớp với kết quả mô phỏng phần mềm R. Điều đó khẳng định tính chuẩn xác của các thuật toán trong hệ thống. Đặc biệt, với ngân hàng câu hỏi chất lượng, hệ thống trắc nghiệm thích ứng cho phép rút ngắn thời gian làm bài của thí sinh mà vẫn đạt được độ chuẩn xác. Giáo viên có thể dễ dàng quản lý ngân hàng câu hỏi, quản lý thí sinh và thu thập kết quả thi của thí sinh một cách tự động, dễ dàng theo dõi sự tiến bộ của HS và có những trợ giúp kịp thời. Nghiên cứu đã tạo tiền đề để phát triển hệ thống DG thích ứng kết hợp dạy học phân hoá cho người học không chỉ ở môn Ngữ văn mà các môn học khác nếu các nhóm NL hoặc các kỹ năng liên quan được xác định rõ ràng, có thể DG được.

Tài liệu tham khảo

- [1] Sái Công Hồng, Lê Thái Hưng, Lê Thị Hoàng Hà, Lê Đức Ngọc (2017), *Giáo trình Kiểm tra đánh giá trong dạy học*, NXB Đại học Quốc gia Hà Nội.
- [2] A. Primacop, (1976), *Phương pháp đọc sách*, NXB Giáo dục, Hà Nội, tr.19.
- [3] Bộ Giáo dục và Đào tạo, (2014), *Tài liệu tập huấn dạy học và kiểm tra, đánh giá kết quả học tập theo định hướng phát triển năng lực học sinh*.
- [4] Nguyễn Hải Châu - Lê Thị Mỹ Hà, (2012), *PISA và các dạng câu hỏi*, NXB Giáo dục Việt Nam.
- [5] Đỗ Thu Hà - Viện Khoa học Giáo dục Việt Nam, *Vận dụng cách thiết kế câu hỏi đánh giá năng lực đọc hiểu của PISA vào môn Ngữ văn*.
- [6] Nguyễn Thái Hòa, (2004), *Vấn đề đọc hiểu và dạy đọc hiểu*, Tạp chí Thông tin Khoa học Sư phạm, số 8.
- [7] Nguyễn Thanh Hùng, (2017), *Kỹ năng đọc hiểu Văn*, NXB Đại học Sư phạm, Hà Nội.
- [8] Alison Wolf, (1995), *Competence-Based Assessment*.
- [9] Eric Witty - Barbara Gaston, (2008), *Competency Based Learning and Assessment, ETITO*.
- [10] *Organization for Economic Cooperation and Development*, (2002).
- [11] Rod Powers - Jennifer Lawler, (2007), *ASVAB For Dummies*, John Wiley @Sons Published house.
- [12] Reckase, M. D, (2003), *Item pool design for computerized adaptive tests*, Paper presented at annual meeting of the National Council on Measurement in Education, Chicago, IL.
- [13] *Quebec Educational Reform*, (2005).
- [14] *Singapore Workforce Development Agency – Quality Assurance Division*, (2012), *Develop competency – based assessment, plans version 1*.
- [15] Weiss, D. J. & Kingsbury, G. G, (1984), *Application of computerized adaptive testing to educational problems* Journal of Educational Measurement, 21, p.361-375.

DEVELOPING COMPUTERIZED ADAPTIVE TESTING TO ASSESS THE READING COMPREHENSION COMPETENCE OF 10TH GRADERS

Le Thai Hung¹, Tran Thi Hoa²,
Dang Thi May³, Hoang Lan Huong⁴

¹ Email: hunglethai82@gmail.com

² Email: tranhoak56@gmail.com

³ Email: dangmaykhd@gmail.com

⁴ Email: huonghoangsp@gmail.com

VNU University of Education,
Vietnam National University, Hanoi
144 Xuan Thuy, Cau Giay, Hanoi, Vietnam

ABSTRACT: *Computerized Adaptive Testing (CAT) is a form of assessment test which requires fewer test questions to arrive at precise measurements of examinees' competence. One of the core technical components in building a CAT is mathematical algorithms which estimate examinee's ability and select the most appropriate test questions for the estimation. Those mathematical algorithms serve as a locomotive in operating the system of adaptive multiple-choice questions on computers. Our research aims to develop the essential mathematical algorithms to a computerised system of adaptive multiple-choice tests. We also build a question bank of 500 multiple-choice questions standardised by IRT theory with the difficulty level follows the normal distribution satisfying Kolmogorov-Smirnov test to measure the reading comprehension competence of students in grade 10. The initial outcome of our experiment on the question bank shows that the question bank satisfies the requirements from a psychometric model and the constructed mathematical algorithms meets the criteria to apply in computerised adaptive testing.*

KEYWORDS: Computerized Adaptive Testing; competence measurement; reading comprehension competence.