Đánh giá chất lượng ngân hàng đề thi trắc nghiệm khách quan môn Nhân học đại cương bằng mô hình RASCH và phần mềm QUEST

Bùi Ngọc Quang

***Tóm tắt*— Bài viết trình bày (1) tổng quan nghiên cứu về lịch sử hình thành phương pháp trắc nghiệm khách quan với sự phát triển của khoa học đo lường và đánh giá kết quả học tập của người học qua phương pháp này; (2) vận dụng lý thuyết khảo thí cổ điển và khảo thí hiện đại vào việc phân tích, đánh giá chất lượng ngân hàng đề thi trắc nghiệm môn Nhân học đại cương dựa trên mô hình RASCH và phần mềm QUEST qua việc xác định độ khó của câu hỏi thi, chất lượng của các phương án sai, độ phân biệt của câu hỏi thi, hệ số tương quan giữa điểm của câu hỏi thi với điểm toàn bài, xác suất khả năng mỗi phương án trả lời được lựa chọn, thang đo năng lực của thí sinh, "ngưỡng" độ khó cho một câu hỏi trắc nghiệm, sai số tính toán, độ tin cậy của đề thi… và qua đó (3) đề xuất một số giải pháp, hướng đến việc áp dụng tối ưu phương pháp trắc nghiệm khách quan tại Trường Đại học Khoa học Xã hội và Nhân văn, Đại học Quốc gia TP. Hồ Chí Minh.**

***Từ khóa*—đánh giá, ngân hàng đề thi, trắc nghiệm khách quan, RASCH, QUEST.**

1 TỔNG QUAN NGHIÊN CỨU

Phương pháp trắc nghiệm khách quan (TNKQ) xuất hiện từ thế kỷ thứ 19 do nhà khoa học người Anh Francis Galton nghĩ ra để đo trí thông minh của con người. Năm 1904, Alfred Binet – nhà tâm lý học người Pháp, đã xây dựng các bài trắc nghiệm để xác định các trẻ em bị khiếm khuyết về mặt tâm thần dẫn đến việc không thể tiếp thu bài học theo cách dạy thông thường ở trường. Năm 1910, trắc nghiệm của Alfred Binet được dịch và sử dụng ở Mỹ. Năm 1920, Edward Thorndike – nhà tâm lý học người Mỹ, đã dùng TNKQ để đo trình độ người học. Sau đó, phương pháp này được phát triển và áp dụng rộng rãi trên toàn thế giới.

Hiện nay, trên thế giới khoa học đánh giá trong giáo dục đang phát triển mạnh mẽ, đặc biệt ở Mỹ cũng như các nước thuộc khối OECD[1]. Lĩnh vực khoa học về đo lường và đánh giá trong giáo dục bắt đầu phát triển và hoàn thiện dần lý thuyết khảo thí cổ điển vào đầu thập niên 1970, sau đó tiếp tục phát triển cho đến ngày nay và trở thành lý thuyết khảo thí hiện đại. Cần ghi nhận trong quá trình phát triển này có sự đóng góp của Ralph Tyler (1949) – một trong những người đầu tiên đưa ra khái niệm đo lường, đánh giá. Quan điểm của ông về vai trò của đánh giá trong giáo dục đã góp phần đáng kể cho việc phát triển chương trình đào tạo và đánh giá giáo dục, và là nền tảng lý luận cho việc thực hành đánh giá TNKQ sau này.

Trong số các công trình nghiên cứu công phu về lĩnh vực đánh giá và đo lường trong giáo dục trên thế giới là "Educational Measurement and Evaluation" (Đo lường và đánh giá trong giáo dục) của Jum C. Nunnally (1964) [10]; "Measuring Educational Achievement" (Đo lường thành tích giáo dục) của Robert L. Ebel (1965) [5] và "Constructing Achievement Tests" (Thiết kế các đề thi đánh giá thành tích học tập) của Norman E. Gronlund (1982) [7]; các tác phẩm này mô tả rất chi tiết phương pháp đo lường đánh giá định lượng kết quả học tập của người học. Benjamin S. Bloom, George F. Madaus, và Thomas J. Hastings (1981) [2] với nghiên cứu "Evaluation to improve learning" (Đánh giá để thúc đẩy học tập), viết về kỹ thuật đánh giá kết quả học tập của người học nhằm tư vấn, hỗ trợ người dạy sử dụng việc đánh giá như một công cụ để cải tiến toàn bộ quá trình dạy và học…

Bài nhận ngày 08 tháng 12 năm 2016, hoàn chỉnh sửa chữa ngày 25 tháng 10 năm 2017

Bùi Ngọc Quang - Trường Đại học Khoa học Xã hội và Nhân văn, ĐHQG-HCM (email: ngocquang.info@gmail.com)

[1] Oganization for Economic Co-operation and Development (Tổ chức Hợp tác và Phát triển kinh tế)

Ở Việt Nam, giáo dục được đề cao và được coi là "quốc sách hàng đầu"[2]; nền giáo dục Việt Nam đã có những biến chuyển tích cực hướng đến nền khoa học và kỹ thuật giáo dục tân tiến của thế giới. Gần đây, vấn đề đo lường và đánh giá trong giáo dục, nói chung và đánh giá kết quả học tập của người học nói riêng nhận được sự quan tâm đặc biệt của Bộ Giáo dục và Đào tạo. TNKQ xuất hiện ở miền Bắc từ những năm 1960. Giai đoạn 1956-1960, các trường ở miền Nam đã sử dụng rộng rãi các hình thức thi trắc nghiệm ở bậc trung học. Sau năm 1975, một số trường đã áp dụng TNKQ song do có những ý kiến trái chiều nên hình thức này lại không được sử dụng. Cho đến gần đây, vấn đề đánh giá giáo dục và trắc nghiệm kết quả học tập mới nhận được sự quan tâm đặc biệt của Bộ Giáo dục và Đào tạo. Một số trường đại học đã bắt đầu xây dựng ngân hàng đề thi trắc nghiệm cho nhiều môn học phổ biến. Năm 2006, Bộ Giáo dục và Đào tạo tổ chức thi TNKQ cho môn Ngoại ngữ và từ năm 2007 tăng thêm các môn Vật lý, Hóa học và Sinh học trong các kỳ thi tốt nghiệp trung học phổ thông và đại học.

Việc đổi mới căn bản hình thức và phương pháp thi, kiểm tra và đánh giá kết quả giáo dục, đào tạo, bảo đảm trung thực, khách quan theo đúng tinh thần Nghị quyết Hội nghị trung ương 8 khóa XI về "đổi mới căn bản, toàn diện giáo dục và đào tạo"[3] qua sự kiện quan trọng của ngành giáo dục là tổ chức kỳ thi trung học phổ thông quốc gia vào năm 2015. Đây là kỳ thi 2 trong 1, được gộp bởi hai kỳ thi là kỳ thi tốt nghiệp trung học phổ thông và kỳ thi tuyển sinh đại học và cao đẳng. Trong kỳ thi trung học phổ thông quốc gia năm 2017, các môn Toán, Khoa học tự nhiên (Vật lý, Hóa học, Sinh học), Khoa học xã hội (Lịch sử, Địa lý, Giáo dục công dân), Ngoại ngữ đều thi theo hình thức trắc nghiệm.

Có nhiều nhà giáo dục đã nghiên cứu về trắc nghiệm và đo lường kết quả học tập như Lâm Quang Thiệp (1994) [8] với "Những cơ sở của kỹ thuật trắc nghiệm"; Dương Thiệu Tống (1995) [3] với "Trắc nghiệm và đo lường thành quả học tập"; Lý Minh Tiên (2004) [9] với "Kiểm tra và đánh giá thành quả học tập của học sinh bằng trắc nghiệm khách quan"; Phạm Xuân Thanh (2011) [12] đã giới thiệu và vận dụng mô hình RASCH và phần mềm QUEST vào việc phân tích và đánh giá chất lượng các câu hỏi/ đề thi trắc nghiệm khách quan trong các kỳ thi đại học, trung học phổ thông… Các nghiên cứu này đều đã trình bày một cái nhìn tổng quan về đo lường và đánh giá trong giáo dục, các phương pháp trắc nghiệm, đánh giá kết quả học tập, và việc ứng dụng, áp dụng khoa học đo lường và đánh giá trong giáo dục trên thế giới và Việt Nam vào thực tiễn… Đây là những tài liệu hữu ích cho giảng viên (GV), cán bộ quản lý giáo dục và những người có quan tâm, nghiên cứu việc đánh giá kết quả học tập của người học.

2 KẾT QUẢ NGHIÊN CỨU

2.1 Thông tin chung về kết quả thi

Bộ đề thi TNKQ môn Nhân học đại cương của Trường Đại học Khoa học Xã hội và Nhân văn, Đại học Quốc gia TP. Hồ Chí Minh (Trường ĐH KHXH&NV, ĐHQG-HCM), gồm 3 đề thi với 70 câu hỏi TNKQ; vị trí của câu hỏi và đáp án được thay đổi tuỳ vào mỗi đề thi. Mỗi đề thi gồm 70 câu hỏi, từ câu 1 đến câu 70, với loại trắc nghiệm nhiều lựa chọn (MCQs: Multiple-Choice Questions) và đảm bảo gần hết các bước kỹ thuật xây dựng câu TNKQ và cũng đảm bảo các mức độ nhận thức theo thang nhận thức của Bloom, nhưng chỉ gồm 3 mức độ biết, hiểu, và vận dụng ở mức độ thấp nhất.

Học kỳ I, năm học 2015-2016 đã sử dụng 03 đề thi (gồm mã đề 001, mã đề 002, mã đề 003) bằng cách bốc thăm ngẫu nhiên từ 300 câu hỏi có sẵn. Thời gian thi là 75 phút; mỗi phòng thi sử dụng cả 3 mã đề thi và phát đề thi xen kẽ theo chỗ ngồi của sinh viên (SV) để tránh tình trạng tham khảo đáp án của nhau.

Trong giới hạn của đề tài nghiên cứu khoa học mà kết quả của nó được trình bày trong bài viết này, nhóm tác giả chỉ phân tích đề thi và kết quả thi của mã đề thi số 002 với dữ liệu gốc của mã đề thi này gồm có 71 biến, bao gồm: MSSV là mã số SV và C1-C70 là kết quả trả lời của 70 câu hỏi trắc nghiệm trong tổng số 277 SV tham gia. Thông tin chung về kết quả thi được thống kê như sau:

[2] lần đầu tiên được quy định tại Điều 35, Hiến pháp năm 1992
[3] Nghị quyết số 29-NQ/TW ngày 4 tháng 11 năm 2013 của Ban Chấp hành Trung ương

BẢNG 1
THỐNG KÊ ĐIỂM THI CỦA SINH VIÊN

Điểm	< 5,0	5,0 – 6,5	7,0 – 8,5	> 8,5
Xếp loại	Không đạt	Trung bình	Khá	Giỏi
Số lượng	6	127	130	14
Tỷ lệ (%)	2,17	45,85	46,93	5,05

Số liệu thống kê trong *Bảng 1* cho thấy số lượng thí sinh có điểm thi toàn bài trên 5 điểm là khá cao, chiếm 97,83%); chỉ có 2,17% tương đương với 6 SV có điểm dưới trung bình (điểm < 5,0) và phải học lại; tỷ lệ SV đạt điểm *trung bình* tương đương với tỷ lệ xếp loại *khá* (đều chiếm khoảng 1/2 tổng số thí sinh tham gia thi kết thúc học phần); số thí sinh có tổng điểm thi đạt trên 8,5 điểm chiếm tỷ lệ khá khiêm tốn (5,05%, 14 SV); và không có SV nào đạt điểm tuyệt đối 10/10 (tổng số câu trả lời đúng cao nhất của SV là 67/70 câu hỏi).

2.2 Sự phù hợp của câu hỏi thi

2.2.1Mức độ phù hợp với mô hình RASCH

Khi dữ liệu kết quả thi phù hợp với mô hình RASCH [6], [12] thì trị số kỳ vọng của các bình phương trung bình (Mean Square) xấp xỉ bằng 1 và trị số kỳ vọng t xấp xỉ bằng 0 (nghĩa là Mean phải bằng hoặc gần 0; và độ lệch chuẩn SD phải bằng hoặc gần bằng 1).

Các số liệu về giá trị trung bình Mean và độ lệch chuẩn SD có được khi xử lý dữ liệu kết quả thi bằng phần mềm QUEST [1], [12] cho thấy dữ liệu dùng để phân tích trong Bảng 2 là phù hợp với mô hình RASCH.

BẢNG 2
DỮ LIỆU PHÂN TÍCH TRONG MÔ HÌNH RASCH

				Khi dữ liệu phù hợp với mô hình thì:
Summary of item Estimates				
==========				
Mean		0		Mean phải bằng hoặc gần 0
SD		1,11		SD phải bằng hoặc gần 1
SD (adjusted)		1,09		
Reliability of estimate		0,98		
Fit Statistics				
==========				
Infit Mean Square		Outfit Mean Square		
Mean	1	Mean	0,97	Mean phải bằng hoặc gần 1
SD	0,07	SD	0,14	SD phải bằng hoặc gần 0
Summary of case Estimates				
==========				
Mean		0,98		
SD		0,62		
SD (adjusted)		0,54		
Reliability of estimate		0,76		
Fit Statistics				
==========				
Infit Mean Square		Outfit Mean Square		
Mean	1	Mean	0,97	Mean phải bằng hoặc gần 1
SD	0,10	SD	0,22	SD phải bằng hoặc gần 0

2.2.2Mức độ phù hợp của các câu hỏi thi

Trong biểu đồ Item Fit qua Bảng 3 dưới đây, mỗi câu trắc nghiệm được biểu thị bằng dấu *, các câu trắc nghiệm nằm trong 2 đường chấm thẳng đứng có giá trị trung bình bình phương độ phù hợp INFIT MNSQ nằm trong giới hạn [0,77; 1,30] sẽ phù hợp với mô hình RASCH, nếu câu trắc nghiệm nào không phù hợp thì loại bỏ.

BẢNG 3
BIỂU ĐỒ VỀ SỰ PHÙ HỢP CỦA CÁC CÂU HỎI THI

```
------------------------------------------------------------------------------
Item Fit                                                    20/ 4/16 19: 9
all on dulieu (N = 277 L = 70 Probability Level= 0,50)
------------------------------------------------------------------------------
INFIT
 MNSQ     0,56      0,63      0,71      0,83      1,00      1,20      1,40
-----------+---------+---------+---------+---------+---------+---------+------
  1 item 1                          .              *              .
  2 item 2                          .              |    *         .
  3 item 3                          .              *              .
  4 item 4                          .              *              .
  5 item 5                          .       *      |              .
  6 item 6                          .              |*             .
  7 item 7                          .              *              .
  8 item 8                          .              |*             .
  9 item 9                          .              | *            .
 10 item 10                         .             * |             .
 11 item 11                         .              |*             .
 12 item 12                         .             *|              .
 13 item 13                         .             *|              .
 14 item 14                         .           *  |              .
 15 item 15                         .             * |             .
 16 item 16                         .              |*             .
 17 item 17                         .              *              .
 18 item 18                         .              | *            .
 19 item 19                         .              |             *.
 20 item 20                         .            *  |             .
 21 item 21                         .           *   |             .
 22 item 22                         .               *             .
 23 item 23                         .              | *            .
 24 item 24                         .              |  *           .
 25 item 25                         .       *      |              .
 26 item 26                         .              |*             .
 27 item 27                         .              |  *           .
 28 item 28                         .              |  *           .
 30 item 30                         .              |*             .
 31 item 31                         .              *              .
 32 item 32                         .              |     *        .
 33 item 33                         .           *  |              .
 34 item 34                         .            *  |             .
 35 item 35                         .            *  |             .
 36 item 36                         .               *             .
 37 item 37                         .           *  |              .
 38 item 38                         .            *  |             .
 39 item 39                         .              *|             .
 40 item 40                         .             * |             .
 41 item 41                         .              |  *           .
 42 item 42                         .              |*             .
 43 item 43                         .              |*             .
 44 item 44                         .             * |             .
 45 item 45                         .           *   |             .
 46 item 46                         .              *|             .
 47 item 47                         .           *   |             .
 48 item 48                         .               *             .
 49 item 49                         .              |       *      .
 50 item 50                         .        *     |              .
 51 item 51                         .              | *            .
 52 item 52                         .              *|             .
 53 item 53                         .             * |             .
 54 item 54                         .              |  *           .
```

```
 55 item 55                    .                *              .
 56 item 56                    .          *     |              .
 57 item 57                    .             * |              .
 58 item 58                    .               |  *           .
 59 item 59                    .               |        *     .
 60 item 60                    .             * |              .
 61 item 61                    .               | *            .
 62 item 62                    .               |  *           .
 63 item 63                    .              *|              .
 64 item 64                    .               *              .
 65 item 65                    .               *              .
 66 item 66                    .            *  |              .
 67 item 67                    .            *  |              .
 68 item 68                    .            *  |              .
 69 item 69                    .               |   *          .
 70 item 70                    .               *
```

Biểu đồ trên cho thấy các câu hỏi đều có chỉ số INFIT MNSQ nằm trong giới hạn [0,77; 1,30] nên đều phù hợp với mô hình RASCH, ngoại trừ câu C29 đã được loại ra khỏi mô hình này do có giá trị INFIT MNSQ nằm ngoài giới hạn cho phép nêu trên.

2.3 Phân bố độ khó câu hỏi thi và năng lực thí sinh

Các thông tin về kết quả tính toán năng lực của thí sinh (case estimate) cho thấy năng lực trung bình của mẫu thí sinh tham gia làm bài thi trắc nghiệm là (θ_{tb} =0,98), lớn hơn so với độ khó chung của đề thi (σ_{tb} =0).

Biểu đồ phân bố độ khó câu hỏi kiểm tra và năng lực thí sinh cho thấy mức độ phù hợp của đề kiểm tra đối với thí sinh dự kiểm tra. Khi xử lý bằng phần mềm QUEST sẽ cho một biểu đồ phân bố năng lực SV và độ khó của các câu hỏi trong đề kiểm tra.

BẢNG 4
MA TRẬN NĂNG LỰC THÍ SINH VÀ ĐỘ KHÓ CỦA CÂU HỎI THI

```
------------------------------------------------------------------------------
Item Estimates (Thresholds)                                  20/ 4/16 19: 9
all on dulieu (N = 277 L = 70 Probability Level= 0,50)
------------------------------------------------------------------------------
  4,0                                |
          NĂNG LỰC CAO               |        RẤT KHÓ
                                     |
                                     |
                               X     |
                                     |
                               X     |
  3,0                                |
                                     |
                               X     |
                                     |
                              XX     |
                              XX     |   32
                             XXX     |
  2,0                       XXXX     |
                            XXXX     |   20
                         XXXXXXXX    |
                            XXXX     |   44
                        XXXXXXXXX    |   18    34
                          XXXXXXX    |    9
                    XXXXXXXXXXXXXX   |   22
  1,0     XXXXXXXXXXXXXXXXXXXXX      |    2     4    19    24    25   37   43   64
                XXXXXXXXXXXXXXX      |   14    31    49    51
                        XXXXXX       |   21
                    XXXXXXXXXXXXXX   |   30    45    56    60    62
```

```
                  XXXXXXXXXXXX    |   10    17    23    59
                          XXXX    |    3    15    26    41    68
  0                     XXXXXX    |    5    27    36    46    54    58    70
                            XX    |   50    53    61    69
                            XX    |
                             X    |    1    11    28    42    47
                                  |
                                  |   55
                                  |   48    52
 -1,0                             |    6
                                  |    8    35    38    39
                                  |   13
                                  |   16    40
                                  |   33
                                  |    7    63
                                  |
 -2,0                             |   66
                                  |   57
                                  |
                                  |
                                  |
                                  |
                                  |   65
 -3,0                             |
                                  |
                                  |
                                  |   12
                                  |
            NĂNG LỰC THẤP         |        RẤT DỄ
                                  |
 -4,0                             |
-------------------------------------------------------------------------------
  Each X represents    2 students
  Some thresholds could not be fitted to the display
===============================================================================
```

Khi phân tích độ khó của câu hỏi thi, phần mềm QUEST cung cấp một biểu đồ dưới dạng ma trận là *Bảng 4* giúp so sánh năng lực của 277 SV với độ khó của 70 câu hỏi thi. Theo biểu đồ ma trận này, các con số bên tay phải cho biết độ khó của các câu hỏi thi còn các dấu X nằm bên trái biểu đồ là sự phân bố năng lực của SV. Mỗi dấu X đại diện cho 2 SV. Nhìn trên biểu đồ có thể thấy rõ nét sự phân bố về độ khó các câu hỏi thi bao trùm hầu hết năng lực của SV: có đến 3/4 số câu hỏi trong đề thi (41 câu) là phù hợp năng lực của SV.

Các câu hỏi có độ khó chỉ đòi hỏi mức năng lực của thí sinh từ -3,35 đến 2,31 (thang Logistic) để có thể hoàn thành bài thi cuối kỳ. Trong khi đó, năng lực thực của SV phân bố từ -0,41 đến 3,48 với trung bình cộng là 0,98 và độ lệch chuẩn là 0,62. Điều này chứng tỏ đề thi có một số câu hỏi dễ hơn nhiều so với năng lực của SV, và chưa có câu hỏi khó để đánh giá những SV có năng lực cao hơn.

Qua biểu đồ ta cũng dễ dàng thấy được có 2 nhóm câu hỏi được chia theo độ khó của câu hỏi so với năng lực của SV. Nhóm thứ nhất là nhóm câu hỏi có độ khó phù hợp với năng lực chung của SV. Nhóm thứ 2 là nhóm có độ khó thấp hơn so với năng lực chung của SV; đây là các câu hỏi dễ, cần được chỉnh sửa hoặc loại bỏ cho phù hợp. Có thể thấy câu hỏi dễ nhất là câu C12, và câu khó nhất là câu C32.

Ngoài ra, biểu đồ phân bố còn cho thấy đề thi còn có những khoảng trống cần được bổ sung bằng một số câu hỏi để đo và phân biệt năng lực của các thí sinh ở nhóm năng lực cao từ trên 2,31 theo thang Logistic (đây là đơn vị dùng để đo ngưỡng độ khó hay năng lực của thí sinh).

2.4 Các chỉ số thống kê của câu hỏi thi

2.4.1Giá trị trung bình bình phương độ hoà hợp

INFIT MNSQ là giá trị trung bình bình phương độ hoà hợp của các câu hỏi thi, những câu hỏi có giá trị này nằm trong khoảng [0,77; 1,30] là phù hợp với mô hình RASCH.

Qua Bảng 5 dưới đây, ta thấy chỉ số INFIT MNSQ của các câu hỏi có giá trị rải từ 0,87 đến 1,27 đều nằm trong khoảng cho phép [0,77; 1,30] nên các câu hỏi trắc nghiệm trong đề thi số 002 là phù hợp với mô hình RASCH; ngoại trừ câu C29 đã được loại ra khỏi mô hình này, do có giá trị INFIT MNSQ = 0.

BẢNG 5
THỐNG KÊ CHỈ SỐ INFIT MNSQ CỦA CÁC CÂU HỎI THI

Câu hỏi	INFIT MNSQ	Câu hỏi	INFIT MNSQ	Câu hỏi	INFIT MNSQ	Câu hỏi	INFIT MNSQ	Câu hỏi	INFIT MNSQ
C1	1,00	C15	0,96	C29	0,00	C43	1,02	C57	0,96
C2	1,10	C16	1,02	C30	1,01	C44	0,96	C58	1,06
C3	1,00	C17	1,01	C31	0,99	C45	0,93	C59	1,16
C4	0,99	C18	1,05	C32	1,12	C46	0,98	C60	0,96
C5	0,88	C19	1,27	C33	0,93	C47	0,93	C61	1,03
C6	1,02	C20	0,95	C34	0,95	C48	1,00	C62	1,07
C7	1,00	C21	0,92	C35	0,94	C49	1,14	C63	0,98
C8	1,02	C22	0,99	C36	1,01	C50	0,90	C64	1,00
C9	1,04	C23	1,05	C37	0,93	C51	1,04	C65	1,00
C10	0,97	C24	1,07	C38	0,94	C52	0,98	C66	0,95
C11	1,01	C25	0,87	C39	0,98	C53	0,95	C67	0,95
C12	0,99	C26	1,03	C40	0,97	C54	1,05	C68	0,94
C13	0,97	C27	1,05	C41	1,05	C55	1,01	C69	1,08
C14	0,93	C28	1,07	C42	1,02	C56	0,90	C70	0,99

2.4.2Độ khó của câu hỏi thi

Theo lý thuyết khảo thí cổ điển, độ khó của câu hỏi thi (P) là tỷ lệ thí sinh trả lời đúng so với tổng số thí sinh tham gia trả lời câu hỏi đó, được sử dụng rộng rãi đối với các câu hỏi đúng/ sai, đa lựa chọn. Theo Osterlind (1989) [11], giá trị độ khó P càng lớn thì câu hỏi càng dễ; và độ khó của câu hỏi nên nằm trong khoảng từ 0,4 đến 0,8.

BẢNG 6
THỐNG KÊ ĐỘ KHÓ CỦA CÂU HỎI THEO LÝ THUYẾT KHẢO THÍ CỔ ĐIỂN

Độ khó P	Mức độ	Số câu	Tỷ lệ %
$P \geq 0,8$	dễ	20	28,6
$0,6 \leq P < 0,8$	trung bình	28	40,0
$0,4 \leq P < 0,6$	tương đối khó	19	27,1
$0,2 \leq P < 0,4$	khó	3	4,3
$P < 0,2$	rất khó	0	0

Trong *Bảng 6* có 20 câu hỏi dễ (chiếm 28,6%), 28 câu hỏi trung bình (chiếm 40%), 19 câu hỏi tương đối khó (chiếm 27,1%), và 3 câu hỏi khó (chiếm 4,3%); không có câu hỏi nào là rất khó.

Áp dụng lý thuyết khảo thí hiện đại, năng lực của SV và độ khó của câu hỏi thi được đánh giá bằng thang Logistic. Theo *Bảng 3*. *Biểu đồ về sự phù hợp của các câu hỏi thi*, các câu hỏi có độ khó trong khoảng [-3,35; 2,31] (theo thang đo Logistic); trong khi đó, năng lực của thí sinh phân bố trong khoảng [-0,41; 3,48] với trung bình cộng là 0,98 và độ lệch chuẩn 0,62. Điều này đòi hỏi phải giảm các câu hỏi quá dễ và tăng một số câu hỏi khó để đo được toàn bộ năng lực của SV.

2.4.3Khả năng nhầm đáp án

Giá trị độ khó P của câu hỏi còn có một thuộc tính nữa: giúp xác định những câu hỏi bị nhầm đáp án. Việc nhầm đáp án là một hiện tượng khá phổ biến trong quá trình thiết kế và xây dựng bộ đề thi TNKQ nhiều lựa chọn. Trong nhiều trường hợp, các nhầm lẫn này là có thể hiểu được: sự đơn điệu trong việc viết câu hỏi TNKQ có thể khiến các chuyên gia thiếu tập trung, dẫn đến thiết kế nhầm đáp án; sự mơ hồ, thiếu rõ ràng trong cách diễn đạt câu hỏi thi có thể gây khó cho thí sinh khi phải xác định phương án trả lời đúng; sự phức tạp về nội dung hoặc thuật ngữ trong các câu hỏi đánh giá các kỹ năng của quá trình nhận thức phức tạp cũng có thể dẫn đến việc xác định phương án trả lời sai.

Những câu hỏi thi TNKQ nhiều lựa chọn bị nhầm đáp án có thể được phát hiện khi người soạn

câu hỏi xem bảng giá trị P và thấy có sự khác biệt lớn giữa dự định và thực tế trả lời của SV.

BẢNG 7
HIỆN TƯỢNG NHẦM ĐÁP ÁN CỦA CÁC CÂU HỎI

Câu hỏi	Đáp án	Phương án chọn				Bỏ sót	Độ khó P	Nhận xét
		A	B	C	D			
C20	A	82	8	8	178	1	0,30	khó
C32	C	131	54	61	31	0	0,22	khó
C44	C	14	110	106	47	0	0,38	khó

Kết quả của *Bảng 7* cho thấy chỉ có 3 trường hợp có khả năng nhầm đáp án có thể xảy ra là ở các câu được ký hiệu là C20, C32 và C44.

2.4.4Chất lượng của các phương án sai

Phương án gây nhiễu, còn gọi là mồi nhử, là các phương án ngoài đáp án. Mồi nhử tốt là mồi nhử có tỷ lệ lựa chọn gần với tỷ lệ mong muốn được tính theo công thức:

$$i = \frac{1-P}{k-1} x100\,\%$$

Trong đó, i: tỷ lệ mồi nhử mong muốn;

P: độ khó của câu hỏi;

k: tổng số phương án trả lời của câu hỏi.

Xét câu hỏi C20 (với 4 lựa chọn) ta có độ khó: P = 0,30 và k = 4 thì tỷ lệ mồi nhử mong muốn là i = 23,33% cho mỗi phương án.

Cách tính này cho phép xác định mồi nhử không hấp dẫn khi tỷ lệ lựa chọn nhỏ hơn 50% tỷ lệ mong muốn.

Câu hỏi thi tốt thường có xác suất lựa chọn các phương án sai (mồi nhử) là tương đương nhau. Các phương án bị bỏ qua hoặc chỉ có một số ít thí sinh lựa chọn chứng tỏ rằng phương án sai đó là quá lộ liễu, làm tăng khả năng đoán đúng của thí sinh. Những phương án sai nhưng thu hút được nhiều thí sinh lựa chọn chắc chắn là những phương án thiên về đánh lừa thí sinh. Các phương án này đều phải được chỉnh sửa. Xét câu C29, các phương án A, B, C đều là các phương án sai quá lộ liễu: tất cả 3 phương án này đều có 0% thí sinh lựa chọn, chứng tỏ mồi nhử của câu hỏi này kém, và cần phải được chỉnh sửa hoặc loại bỏ ngay. Tương tự như vậy, các câu ký hiệu C6, C12, C34 và C40 đều là các câu có mồi nhử kém.

C29. Hành động nào không có trong tín ngưỡng thờ cúng tổ tiên của người Việt?

A. Thờ cúng linh hồn người thân đã mất
B. Cúng giỗ hàng năm
C. Tảo mộ hàng năm
D. Đọc tên những người đã mất trong gia đình trước khi đi ngủ

Đối với những câu dễ (có P ≥ 0,80) thì các phương án nhiễu hầu như ít có tác dụng để đánh giá kiến thức của SV.

2.4.5Độ phân biệt của câu hỏi thi

Độ phân biệt của câu hỏi thi (I) là khả năng mà câu trắc nghiệm phân loại được thí sinh thành những nhóm có trình độ khác nhau trong lĩnh vực mà bài trắc nghiệm cần đo lường. Sự phân biệt này mô tả chi tiết số người trả lời đúng (nằm ở nhóm người đạt điểm cao ở toàn bài) so với số người trả lời sai (nằm ở nhóm người đạt điểm thấp toàn bài). Công thức tính độ phân biệt của câu hỏi thi là:

$$I = \frac{G_t - G_d}{g}$$

Trong đó, G_t: số SV trả lời đúng ở nhóm cao;

G_d: số SV trả lời đúng ở nhóm thấp;

g: số SV nhóm cao điểm hoặc thấp điểm ở bài trắc nghiệm (chiếm khoảng 27% tổng số SV).

Theo Ebel (1965) [5], các câu hỏi của bài thi nên có chỉ số phân biệt bằng 0,30 hoặc cao hơn. Tuy nhiên, cũng có nhiều người cho rằng độ phân biệt nên nằm trong khoảng chấp nhận từ 0,15 – 0,75. Giá trị độ phân biệt biến thiên trong khoảng (-1, +1), nếu câu hỏi thi có chỉ số phân biệt nhỏ hơn hoặc bằng 0 cần bị loại bỏ hoặc điều chỉnh. Trong các kỳ thi có quy mô lớn, việc sử dụng một số câu hỏi quá dễ hoặc quá khó sẽ dẫn đến độ phân biệt của câu hỏi có thể có giá trị quá thấp hoặc quá cao (độ phân biệt không tốt).

BẢNG 8
THỐNG KÊ MỨC ĐỘ PHÂN BIỆT GIỮA CÁC CÂU HỎI THI

Điều kiện	Số câu	Mức độ	Câu hỏi thi
0,35 ≤ I ≤ 0,75	17	Xuất sắc	C4, C5, C10, C14, C15, C20, C21, C25, C34, C37, C44, C45, C46, C50, C56, C60
0,25 ≤ I < 0,35	14	Tốt	C3, C11, C17, C22, C30, C31, C35, C38, C43, C47, C51, C53, C64, C68
0,15 ≤ I < 0,25	21	Tạm được	C1, C6, C9, C13, C18, C23, C24, C26, C27, C33, C36, C39, C40, C42, C48, C52, C55, C58, C61, C62, C69
I < 0,15	18	Kém	C2, C7, C8, C12, C16, C19, C28, C29, C32, C41, C49, C54, C57, C59, C63, C65, C66, C70

Kết quả phân tích dữ liệu cho thấy độ phân biệt rải từ -0,17 đến 0,68 và có 52 câu có độ phân biệt từ 0,15 – 0,75, nằm trong khoảng chấp nhận được (chiếm 74,3% tổng số câu hỏi trong đề thi); có độ phân biệt < 0,15 và vì vậy mà cần phải được chỉnh sửa trước khi đưa vào ngân hàng câu hỏi là 18 câu hỏi được ký hiệu là C2, C7, C8, C12, C16, C19, C28, C29, C32, C41, C49, C54, C57, C59, C63, C65, C66, và C70.

2.4.6Hệ số tương quan giữa điểm của câu hỏi thi với điểm toàn bài

Giữa kết quả điểm của từng câu hỏi thi với điểm chung của toàn bài thi phải có mối tương quan dương. Việc này có thể kiểm tra dễ dàng bằng các hàm trong Excel hoặc SPSS, QUEST, hoặc tính theo công thức sau:

$$r_{pbis} = \frac{(\overline{x_i} - \overline{x_c})}{\sigma_c}\sqrt{\frac{p_i}{q_i}}$$

Trong đó, $\overline{x_i}$: trung bình cộng điểm của người trả lời được câu hỏi *i* đang xem xét mối tương quan với bài thi;

$\overline{x_c}$: trung bình cộng điểm của toàn bài thi;

p_i : tỷ lệ người trả lời đúng câu hỏi *i*, (độ khó của câu hỏi *i*);

q_i : tỷ lệ người trả lời sai câu hỏi *i*, ($q_i = 1 – p_i$);

σ_c : độ lệch chuẩn của điểm cả bài thi.

Mối tương quan chặt chẽ giữa câu hỏi thi và toàn bài thi góp phần làm tăng độ tin cậy của bài thi. Cần giữ lại những câu hỏi thi có mối tương quan cao và loại bỏ những câu hỏi thi có mối tương quan thấp hoặc dưới 0 để làm tăng độ tin cậy của đề thi.

Giữa kết quả điểm của từng câu hỏi thi với điểm chung của toàn bài thi phải có mối tương quan dương. Theo Dương Thiệu Tống (2000) [4], chúng có mối tương quan giữa 2 biến định lượng như sau:

- 0,8 – 1: tương quan cao đáng tin cậy;
- 0,6 – 0,79: tương quan vừa phải;
- 0,4 – 0,59: tạm được;
- 0,2 – 0,39: tương quan ít;
- 0 – 0,19: tương quan không đáng kể.

BẢNG 9
THỐNG KÊ MỨC ĐỘ TƯƠNG QUAN CỦA CÁC CÂU HỎI THI

Hệ số tương quan	Mức độ	Số câu	Câu hỏi
0,8 - 1,00	tương quan cao	0	
0,6 - 0,79	tương quan vừa phải	0	
0,4 - 0,59	tạm được	6	C14, C21, C50, C56, C5, C25
0,2 - 0,39	tương quan ít	39	C1, C9, C18, C48, C51, C11, C26, C63, C68, C36, C39, C52, C30, C43, C13, C17, C40, C3, C57, C64, C4, C31, C66, C22, C46, C10, C15, C33,C35, C38, C53, C44, C60, C20, C34, C67, C47, C37, C45
0 - 0,19	tương quan không đáng kể, may rủi	23	C29, C70, C49, C32, C28, C65, C2, C16, C6, C7, C8, C12, C27, C58, C62, C24, C41, C54, C23, C55, C61, C69, C42
< 0	tương quan nghịch	2	C19, C59

Bảng 9 cho thấy chỉ có 2 câu là C19 và C59 có hệ số tương quan giữa điểm của câu hỏi thi với điểm toàn bài thi (point-biserial) < 0 (tương ứng là -0,16 và -0,03) nên cần phải loại bỏ để làm tăng độ tin cậy của đề thi; giữa kết quả điểm của từng câu hỏi với điểm chung của toàn bài thi đều là

tương quan thuận nhưng hệ số tương quan này tương đối thấp: chỉ từ 0 đến 0,49.

2.4.7Xác suất khả năng mỗi phương án trả lời được lựa chọn

P-value là giá trị thống kê cho biết hệ số tương quan (Point Biserial) tính toán được là có ý nghĩa thống kê ở mức nào (hay nói cách khác là xác suất khả năng mỗi phương án trả lời được lựa chọn), thông thường phải ≤ 0,05 (có ý nghĩa thống kê ở mức α = 0,05).

Trong số 70 câu hỏi được phân tích trên, có 5 câu có giá trị P-value > 0,05 (gồm C28, C29, C32, C49, và C59) là chưa đạt yêu cầu và cần được xem xét lại vì nó không có ý nghĩa thống kê ở mức α = 0,05. Các câu còn lại đều có P-value ≤ 0,05 là đạt yêu cầu; nghĩa là nó có mức ý nghĩa thống kê ở mức α = 0,05.

2.4.8Thang đo năng lực của thí sinh

Giá trị Mean ability là thang đo năng lực của thí sinh với việc đưa ra sự lựa chọn của mình. Phương án trả lời đúng phải có chỉ số Mean ability cao hơn các phương án trả lời sai. Với kết quả xử lý dữ liệu bằng phần mềm QUEST, thì có 9 câu (xem Bảng 10) cần được xem xét lại vì có Mean ability của phương án đúng nhỏ hơn phương án sai. Các câu còn lại đều có giá trị Mean ability của phương án trả lời đúng lớn hơn Mean ability của phương án trả lời sai.

BẢNG 10
THỐNG KÊ GIÁ TRỊ MEAN ABILITY LỚN HƠN PHƯƠNG ÁN ĐÚNG

Câu hỏi	Phương án trả lời đúng		Phương án trả lời sai	
	Phương án	Mean ability	Phương án	Mean ability
C1	A	1,04	B	1,08
C7	C	1,01	A	1,32
C19	D	0,89	B	1,17
C23	D	1,07	A	1,15
C26	C	1,07	D	1,41
C27	C	1,05	D	1,71
C28	A	1,01	C	1,07
C49	B	1,01	D	1,18
C59	D	0,96	C	1,13

2.4.9"Ngưỡng" độ khó của câu hỏi

Thresholds là "ngưỡng" độ khó cho một câu hỏi trắc nghiệm cũng là mức khả năng, năng lực yêu cầu mà người làm trắc nghiệm phải có để có cơ may 50% trả lời đúng câu hỏi ấy và được biểu thị trên thang đo Logistic. Với 70 câu hỏi này ta thấy chỉ số Thresholds nằm trong khoảng [-3,35; 2,31], trong khi đó ngưỡng năng lực của thí sinh phân bố trong khoảng [-0,41; 3,48]; điều này cho thấy đề thi này có nhiều câu dễ so với năng lực tối thiểu của SV và không có câu hỏi nào quá khó vượt ngưỡng năng lực của SV. Ví dụ, câu C12 có "ngưỡng" độ khó Thresholds = -3,35 là một câu dễ vì nó chỉ đòi hỏi người có ngưỡng khả năng là -3,35 để có cơ may 50% làm đúng câu ấy.

2.4.10Sai số tính toán

Error là sai số tính toán; thông số này cho thấy độ tin cậy của số liệu tính được cho từng câu hỏi thi, thông thường là < 0,2. Kết quả phân tích cho thấy đề thi có 60 câu hỏi có Error < 0,2 và 10 câu hỏi có Error ≥ 2, gồm C7, C12, C13, C16, C33, C40, C57, C63, C65 và C66.

2.4.11Độ tin cậy của đề thi

Độ tin cậy của đề thi (δ) được tính theo nhiều công thức khác nhau. Thường được sử dụng là độ tin cậy được xác định dựa trên tính ổn định bên trong của đề thi. Đề thi được đánh giá tốt khi có độ tin cậy ≥ 0,8.

Kết quả tính toán bằng phần mềm QUEST cho thấy độ tin cậy của đề thi đạt 0,98. Đây là một đề thi có độ tin cậy cao.

3 KẾT LUẬN VÀ KIẾN NGHỊ

3.1 Kết luận

Các phân tích trên đây đã chỉ ra những ưu điểm cũng như tồn tại của các câu hỏi thi trắc nghiệm trong mã đề 002 làm cơ sở cho việc chỉnh sửa và lựa chọn các câu hỏi tốt để đưa vào ngân hàng câu hỏi thi trắc nghiệm môn Nhân học đại cương. Việc phân tích, đánh giá đề thi bằng các phần mềm ứng

dụng là thao tác cần thiết và rất quan trọng trong quá trình xây dựng ngân hàng câu hỏi thi.

Do đây là đề thi đánh giá kết thúc môn học nên việc lựa chọn và sử dụng nhiều câu hỏi dễ, phù hợp với mục tiêu môn học là hoàn toàn có thể chấp nhận được. Tuy nhiên, nếu đây là kỳ thi có mục đích phân hạng cao thấp về năng lực của thí sinh thì đây là đề thi trung bình do khó phân biệt được các nhóm thí sinh khá, giỏi.

Ưu điểm:

- Chất lượng đề thi tương đối tốt;
- Đa số câu hỏi phù hợp với năng lực của thí sinh;
- Đề thi có độ tin cậy cao;
- Độ phân biệt của đề thi chấp nhận được;

Các câu hỏi trong đề thi có độ phù hợp cao, phù hợp với mô hình RASCH.

Hạn chế:

- Có 3 câu hỏi thi có hiện tượng nhầm đáp án, trường hợp này cần đặc biệt lưu ý để rút kinh nghiệm cho công tác soạn câu hỏi thi;
- Đề thi có nhiều câu hỏi dễ so với năng lực trung bình của SV và thiếu những câu hỏi khó để đánh giá SV có năng lực cao (là những SV có mức năng lực từ 2.31 trở lên theo thang Logistic);
- Câu C29 cần được loại bỏ do ngoại lai (100% SV trả lời đúng câu này);
- Chất lượng của các phương án mồi nhử không cao: một số câu có phương án, mồi nhử sai quá lộ liễu và có những phương án thiên về đánh lừa thí sinh. Trong quá trình soạn câu hỏi trắc nghiệm và tổ hợp lại thành đề thi, hay xây dựng ngân hàng câu hỏi thi, cần lưu ý đến chất lượng phương án mồi nhử: nếu chất lượng mồi nhử không đảm bảo sẽ tăng khả năng thí sinh đoán mò hoặc dùng phương pháp loại trừ; do đó, chất lượng câu hỏi thi không đảm bảo sẽ không đánh giá chính xác được năng lực người học.

Đề tài nghiên cứu khoa học mà kết quả của nó được trình bày trong bài viết này đã sử dụng phần mềm QUEST để xử lý và phân tích kết quả thi cuối học kỳ môn Nhân học đại cương trong học kỳ I năm học 2015-2016 dành cho SV chính quy của Trường ĐH KHXH&NV, ĐHQG-HCM một cách hệ thống và rõ ràng.

Việc biên soạn đề thi còn một số hạn chế; kết quả đánh giá chưa khách quan do chưa được xử lý, đánh giá, phân tích và chưa đảm bảo độ tin cậy do nhiều yếu tố khác nhau. Ngoài ra, sau khi GV ra đề và chấm thi môn Nhân học đại cương xong thì hầu như không có công cụ nào để phân tích và xử lý kết quả thi một cách khoa học, chuyên nghiệp nhằm đảm bảo tính khách quan và chất lượng của đề thi. Mặt khác, GV có thể đã được bồi dưỡng việc biên soạn đề thi TNKQ và cách phân tích và xử lý kết quả thi nhưng chưa được thực hành một cách chi tiết, cụ thể, và rõ ràng.

Hy vọng rằng kết quả của đề tài nghiên cứu khoa học này sẽ góp phần giải quyết được các vấn đề bất cập nêu trên.

3.2 Kiến nghị

Từ những kết luận nêu trên, nhóm tác giả thực hiện đề tài nghiên cứu khoa học xin nêu 5 đề xuất – kiến nghị sau đây để nâng cao hiệu quả của việc kiểm tra, đánh giá kết quả học tập của SV Trường ĐH KHXH&NV, ĐHQG-HCM, trong môn Nhân học đại cương nói riêng và toàn bộ các môn học có tổ chức thi trắc nghiệm nói chung:

Thứ nhất, nâng cao nhận thức về việc kiểm tra, đánh giá kết quả học tập cho GV và cả SV: chỉ đạo cho GV các bộ môn tăng cường công tác kiểm tra, đánh giá hơn nữa bằng việc kết hợp linh hoạt các phương pháp trong từng học phần, căn cứ vào mục tiêu, nội dung chương trình để thúc đẩy việc tự học và nghiên cứu của SV nhằm nâng cao năng lực của SV.

Thứ hai, tạo điều kiện cho GV học tập và nghiên cứu sâu lý thuyết đo lường và đánh giá nói chung, lý thuyết khảo thí cổ điển và khảo thí hiện đại nói riêng, và phương pháp biên soạn câu hỏi TNKQ, xây dựng ma trận đề thi, giúp cho đội ngũ GV có kiến thức, kỹ năng và kinh nghiệm để đảm nhận lĩnh vực khoa học mới này; ngoài ra, cũng cần bồi dưỡng cho GV về tin học, ngoại ngữ và việc sử dụng trang thiết bị hiện đại phục vụ cho việc xử lý và phân tích kết quả thi, để kết quả đánh giá có tác dụng với việc dạy và học nhằm nâng cao chất lượng đào tạo chung của Nhà trường.

Thứ ba, Nhà trường nên đầu tư hơn nữa cho GV xây dựng ngân hàng câu hỏi TNKQ, thử nghiệm các đề thi TNKQ một cách nghiêm túc và khoa học; thường xuyên điều chỉnh, bổ sung các câu hỏi mới trong ngân hàng đề thi trắc nghiệm khách quan; công khai hóa quá trình kiểm tra đánh giá kết quả học tập cùng với việc nâng cao chất lượng các phương pháp thi truyền thống để hạn chế, tiến tới chấm dứt việc gian lận trong thi cử.

Thứ tư, bên cạnh việc tổ chức tập huấn, nâng cao trình độ, nghiệp vụ chuyên môn về kiểm tra đánh giá kết quả học tập giúp cho GV nhận biết và hiểu rõ những kiến thức, công thức cơ bản nhất để có thể tự phân tích, đánh giá chất lượng bài thi qua lý thuyết khảo thí cổ điển, Nhà trường cần đầu tư, trang bị cơ sở vật chất, phần mềm chuyên dụng có bản quyền cho việc thiết kế ma trận đề thi, phân

tích, đánh giá chất lượng ngân hàng đề thi TNKQ dựa trên lý thuyết khảo thí hiện đại cho đơn vị chuyên trách là Phòng Khảo thí và Đảm bảo chất lượng; qua đó, sau mỗi đợt thi kết thúc học phần, Phòng Khảo thí và Đảm bảo chất lượng sẽ xử lý dữ liệu bằng phần mềm chuyên dụng và trích xuất kết quả, dữ liệu để cung cấp, thông báo kết quả cho GV ra đề thi những câu hỏi thi có vấn đề cần được chỉnh sửa, điều chỉnh. Điều này sẽ giúp cho Nhà trường tránh lãng phí nhân sự, thời gian, công sức phải tính toán, phân tích dữ liệu thi thủ công như hiện nay.

Thứ năm, Nhà trường cần có chủ trương, quan điểm rõ ràng ở cấp Trường/ cấp Khoa về việc xây dựng, quản lý và sử dụng ngân hàng đề thi/ câu hỏi thi trắc nghiệm; chính thức tuyên truyền cho GV và các đối tượng liên quan về tầm quan trọng và lợi ích của việc xây dựng ngân hàng đề thi chung cho toàn Trường; và xây dựng cơ chế quản lý việc sử dụng ngân hàng đề thi/ câu hỏi thi trắc nghiệm.

TÀI LIỆU THAM KHẢO

[1]. Adams, R. J. & Khoo, S. T. (1996), *QUEST Software*, Camberwell: Quest Software Pty Ltd.

[2]. Bloom, B. S., Madaus, G. F. & Hastings, J. T. (1981), *Evaluation to improve learning*, New York: Mcgraw-Hill.

[3]. Dương Thiệu Tống (1995), *Trắc nghiệm và đo lường thành quả học tập*, TP. Hồ Chí Minh: Trường Đại học Tổng hợp TP. Hồ Chí Minh.

[4]. Dương Thiệu Tống (2000), *Thống kê ứng dụng trong nghiên cứu khoa học giáo dục*, Hà Nội: NXB Đại học Quốc gia Hà Nội.

[5]. Ebel, R. L. (1965), *Measuring Educational Achievement*, Englewood Cliffs: Prentice-Hall.

[6]. Griffin, J. P. (1997), *An introduction to the RASCH model*, Australia: University of Melbourne.

[7]. Gronlund, N. E. (1982), *Constructing achievement tests (3rd ed.)*, Englewood Cliffs: Prentice-Hall.

[8]. Lâm Quang Thiệp (1994), *Những cơ sở của kỹ thuật trắc nghiệm*, Hà Nội: NXB Đại học Quốc gia Hà Nội.

[9]. Lý Minh Tiên (2004), Kiểm tra và đánh giá thành quả học tập của học sinh bằng trắc nghiệm khách quan, Hà Nội: NXB Giáo dục.

[10]. Nunnally, J. C. (1964), *Educational Measurement and Evaluation*, New York: Mc Graw-Hill.

[11]. Osterlind, S. J. (1989), *Constructing test items*, Boston: Kluwer Academic.

[12]. Phạm Xuân Thanh (2011), *Mô hình RASCH và phân tích dữ liệu bằng phần mềm QUEST*, Tài liệu bài giảng khoá đào tạo thạc sĩ Đo lường và đánh giá trong giáo dục, Viện Đảm bảo chất lượng giáo dục, Đại học Quốc gia Hà Nội.

Bùi Ngọc Quang đã nhận bằng thạc sĩ về Đo lường và Đánh giá trong giáo dục từ Viện Đảm bảo Chất lượng Giáo dục, Đại học Quốc gia Hà Nội vào năm 2013. Ông hiện là nghiên cứu sinh chuyên ngành Quản lý Giáo dục tại Trường Đại học Khoa học Xã hội và Nhân văn, ĐHQG-HCM. Từ năm 2009 đến nay, ông là cán bộ chuyên trách công tác đảm bảo chất lượng tại Trường Đại học Khoa học Xã hội và Nhân văn, ĐHQG-HCM. Các mối quan tâm nghiên cứu của ông bao gồm Đo lường và đánh giá trong giáo dục, Quản lý chất lượng trong giáo dục, ICT trong giáo dục.

Evaluation of the quality of multiple choice test bank for the module of Introduction to Anthropology by using the RASCH model and QUEST software

Bui Ngoc Quang
University of Social Sciences and Humanities, VNU-HCM

Abstract—**The paper presents (1) a general view of the history of the development of objective multiple choice testing methods in accordance with the development of measurement science, and the evaluation process of the learners' academic performance by this method; (2) the process of applying classic and modern test theories to analyze and evaluate the quality of multiple choice test bank for the module of Introduction to Anthropology by the RASCH model and QUEST software, which is implemented by the determination of difficulty degree of the questionnaires, the quality of the wrong opinions, the degree of difference among the test questions, the correlation factors between the test score and the whole score, the probability of each option being chosen, the measurement scale for the learners' competence, the "threshold level" of the difficulty level for a multiple choice question, the calculation error, the reliability of the test, etc. and thereby (3) some solutions made towards the optimal application of the objective multiple choice tests at the University of Social Sciences and Humanities, Vietnam National University Ho Chi Minh City.**

Index Terms—**evaluation, test bank, objective test, RASCH, QUEST.**