NGHIÊN CỨU MỨC ĐỘ PHÙ HỢP CỦA VIỆC LƯỢNG GIÁ LÝ THUYẾT BẰNG PHƯƠNG PHÁP TRẮC NGHIỆM KHÁCH QUAN DÀNH CHO SINH VIÊN NĂM THỨ TƯ (HỆ 6 NĂM) TẠI BỘ MÔN NHI TỪ NĂM 2005-2008

Nguyễn Thanh Long
Bộ môn Nhi, Trường Đại học Y Dược Huế

Tóm tắt

Đặt vấn đề: Trắc nghiệm khách quan là một phương pháp lượng giá ngày càng được áp dụng rộng rãi vì có nhiều ưu điểm. Việc đánh giá mức độ phù hợp của đề thi dựa trên phân tích độ khó và độ phân biệt rất cần thiết để có những đề thi tốt. **Phương pháp nghiên cứu:** Đối tượng nghiên cứu là điểm số thi trắc nghiệm khách quan (dạng MCQ) môn lý thuyết nhi của sinh viên Y4 trong 3 năm học (2005-2008). Đánh giá độ khó của đề thi bằng cách so sánh trung bình số học của điểm số với trung bình lý tưởng. Đánh giá độ phân biệt bằng test phân bố chuẩn của điểm số. **Kết quả và kết luận:** Trong năm học 2005-2006: Chỉ một đề thi đạt độ khó phù hợp và hai đề thi đạt độ phân biệt phù hợp. Trong năm học 2006-2007: Không có đề thi nào đạt độ khó phù hợp và chỉ một đề thi đạt độ phân biệt phù hợp. Trong năm học 2007-2008: Không đề thi nào đạt độ khó và độ phân biệt phù hợp.

Từ khóa: Độ khó; độ phân biệt; trắc nghiệm khách quan

Abstract

STUDY ON THE SUITABILITY OF THE PEDIATRIC OBJECTIVE TESTS FOR 4th YEAR STUDENTS (6 YEAR TRAINING PROGRAM) AT PEDIATRIC DEPARTMENT FROM 2005 TO 2008

Nguyen Thanh Long

Background: Objective test with a lot of advantages becomes more and more popular in evaluating students' knowledge. The assessment for the suitability of the question set based on the difficulty level and the differentiating level is very necessary for composing good questions **Methods:** The results of pediatric objective tests (MCQ type) of fourth year students in 3 successive years (2005-2008) were collected. The assessment of difficulty level was based on the comparison between the arithmetic mean and ideal mean. The assessment of differentiating level was based on the test for normal distribution of the marks. **Results and Conclusions:** In school year 2005-2006: There was only one question set having suitable difficulty level and two question sets having suitable differentiating level. In school year 2006-2007: There was no question set having suitable difficulty level and only one question set having suitable differentiating level. In school year 2007-2008: There was no question set having suitable difficulty level or suitable differentiating level.

Key words: difficulty level; differentiating level; objective test

1. ĐẶT VẤN ĐỀ

Trong những năm gần đây trắc nghiệm khách quan là một phương pháp lượng giá được áp dụng phổ biến trong các trường đại học và cao đẳng trên toàn quốc [3]. Ở trường Đại học Y Dược Huế, Bộ Môn Nhi là một đơn vị tiên phong trong việc áp dụng phương pháp này để đánh giá kiến thức lý thuyết của sinh viên. Việc sử dụng phương pháp lượng giá này bước đầu đã giúp hạn chế được một số tiêu cực trong sinh viên và giúp các bộ môn tiết kiệm thời gian chấm thi nhưng cách lượng giá này chỉ phản ánh trung thực trình độ của sinh viên khi các câu hỏi được soạn có độ khó và độ phân biệt phù hợp. Tuy nhiên, chưa được đơn vị nào trong nhà trường thực hiện việc đánh giá chất lượng các đề thi trắc nghiệm khách quan đã được sử dụng. Đề tài này nhằm bước đầu đánh giá mức độ phù hợp của các đề thi trắc nghiệm lý thuyết nhi khoa đã được sử dụng cho đối tượng sinh viên năm thứ 4 trong ba năm học từ 2005 đến 2008. Những kết quả rút ra từ nghiên cứu này sẽ giúp Bộ Môn điều chỉnh lại cách ra đề thi để sự đánh giá bằng trắc nghiệm khách quan có thể phản ánh chân thực nhất trình độ cũng như kết quả học tập của sinh viên.

2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1.Đối tượng nghiên cứu:

Kết quả thi trắc nghiệm khách quan môn lý thuyết Nhi của 3 lớp sinh viên năm thứ 4 (Y4A, Y4B, Y4C) trong 3 năm học liên tiếp (từ 2005 đến 2008).

2.2. Phương pháp nghiên cứu:

2.2.1. Đặc điểm của các đề thi trắc nghiệm khách quan được sử dụng và cách tổ chức thi:

Bộ môn sử dụng bộ câu hỏi nhiều chọn lựa (5 chọn 1) để sinh viên chọn đáp án phù hợp hợp nhất (MCQ)[4]. Đề thi gồm 100 câu hỏi, được hoàn thành trong 60 phút. Mỗi bộ đề thi gồm có 2 đề (A và B) với nội dung giống nhau nhưng được sắp xếp theo thứ tự khác nhau để khắc phục hiện tượng chép kết quả của nhau. Hai sinh viên ngồi gần nhau phải có đề thi khác nhau. Phương án chọn được sinh viên ghi vào phiếu trả lời. Bài thi cũng được phòng giáo vụ rọc phách trước khi giao cho cán bộ bộ môn chấm bằng phiếu đục lỗ. Mỗi bài thi được chấm bởi 2 cán bộ giảng dạy. Những bài có điểm số <5 sẽ được giáo vụ bộ môn chấm lại lần thứ 3.

Thang điểm sử dụng là từ 0 – 10. Trong đó điểm 0 tương ứng với số câu đúng ≤ 20 và điểm 10 tương tứng với số câu đúng ≥ 98.

2.2.2. Tiêu chí đánh giá tính phù hợp của các đề thi trắc nghiệm khách quan:

Hiện nay, để đánh giá chất lượng nói chung của đề thi trắc nghiệm khách quan chúng ta sử dụng nhiều loại tiêu chí. Tuy nhiên hai tiêu chí hàng đầu là độ khó, độ phân biệt của đề thi. Sau đó có thể kể đến các tiêu chí như độ tin cậy, độ giá trị.

- Độ khó của một bài thi trắc nghiệm khách quan được đánh giá bằng cách so sánh điểm số trung bình lý tưởng (TBLT) của bài thi trắc nghiệm khách quan với điểm số trung bình số học (TBSH) của nhóm sinh viên.

Với đề thi sử dụng toàn bộ câu hỏi có nhiều chọn lựa (với 5 phương án chọn 1) thì điểm trung bình lý tưởng sẽ là: TBLT = (100+20)/2 = 60 câu trả lời đúng , tương ứng với 5 điểm.

Trong đó, 100 là số câu trả lời đúng của sinh viên có kiến thức tốt nhất. 20 là số câu trả lời đúng của sinh viên không có kiến thức nhưng chi trả lời ngẫu nhiên.

- Độ phân biệt của đề thi được đánh giá bằng sự phân bố chuẩn (normal distribution) của kết quả thu được. Có nghĩa là đề thi có độ phân biệt tốt nếu kết quả phân bố rải đều ở cả 3 mức độ cao thấp và trung bình với sự ưu thế khu vực trung bình (phân bố hình chuông) [2].

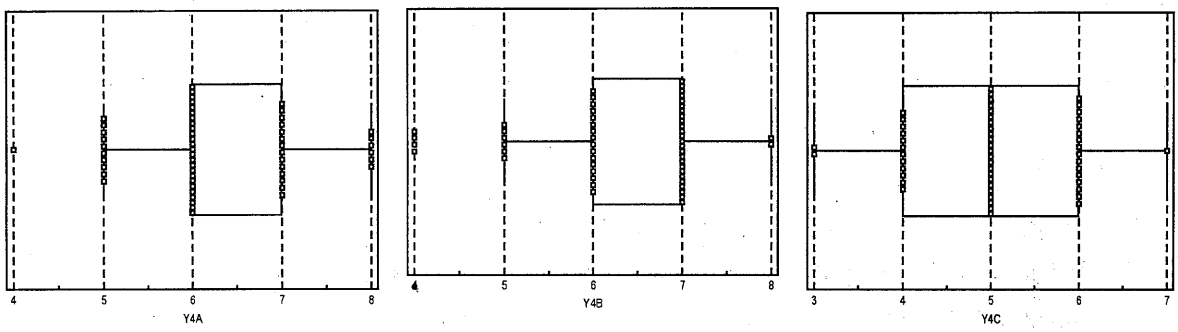
Việc đánh giá phân bố chuẩn của kết quả điểm số được thực hiện bằng phần mềm Medcalc 9.2.1.0 với test phân bố chuẩn D'Agostino-Pearson. Biểu đồ "Box-and-whisker plot" được sử dụng để phân tích phân bố điểm số.

3. KẾT QUẢ

3.1. Tình hình chung về điểm số môn lý thuyết Nhi 1-2 của 3 lớp Y4 qua 3 năm học liên tiếp từ 2005-2008:

Bảng 3.1. Tình hình điểm số môn lý thuyết Nhi 1-2 của 3 lớp Y4 năm học 2005-2006

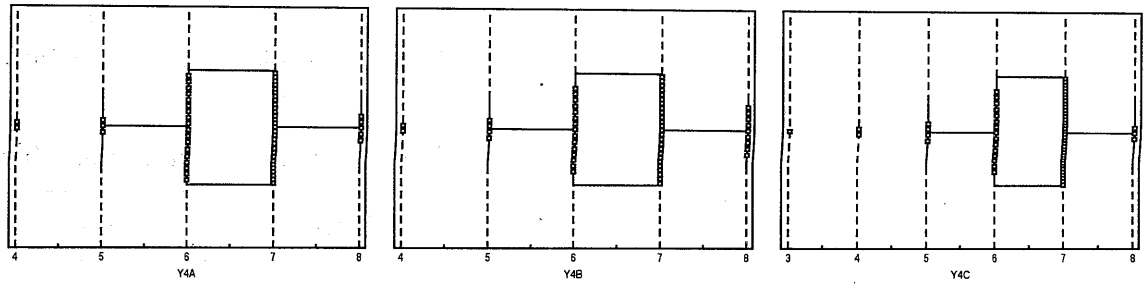
	Y4A	Y4B	Y4C
N	56	53	54
Điểm thấp nhất	4	4	3
Điểm cao nhất	8	8	7
Trung bình số học	6,25	6,28	5,03
Trung vị (Median)	6	7	5
Phương sai (Variance)	0,88	0,97	0,75
Độ lệch chuẩn (SD)	0,93	0,98	0,86
Test phân bố chuẩn	+	-	+



Biểu đồ 3.1. Phân bố điểm số lý thuyết Nhi của ba lớp Y4 năm học 2005-2006

Bảng 3.2. Tình hình chung về điểm số môn lý thuyết Nhi của 3 lớp Y4 năm học 2006-2007

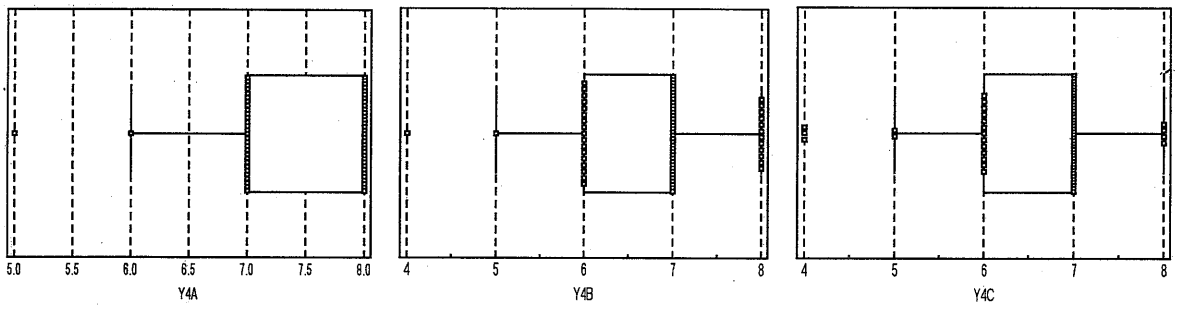
	Y4A	Y4B	Y4C
N	57	55	61
Điểm thấp nhất	4	4	3
Điểm cao nhất	8	8	8
Trung bình số học	6,56	6,63	6,50
Trung vị (Median)	7	7	7
Variance (Phương sai)	0,75	0,93	0,88
Độ lệch chuẩn (SD)	0,86	0,96	0,94
Test phân bố chuẩn	-	+	-



Biểu đồ 3.2. Phân bố điểm số lý thuyết Nhi của ba lớp Y4 năm học 2006-2007

Bảng 3.3. Tình hình chung về điểm số môn lý thuyết Nhi của 3 lớp Y4 năm học 2007-2008

	Y4A	Y4B	Y4C
N	64	67	62
Điểm thấp nhất	5	4	4
Điểm cao nhất	8	8	8
Trung bình số học	7,46	6,85	6,64
Median	8	7	7
Variance	0,38	0,61	0,72
Độ lệch chuẩn (SD)	0,61	0,78	0,85
Test phân bố chuẩn	-	-	-

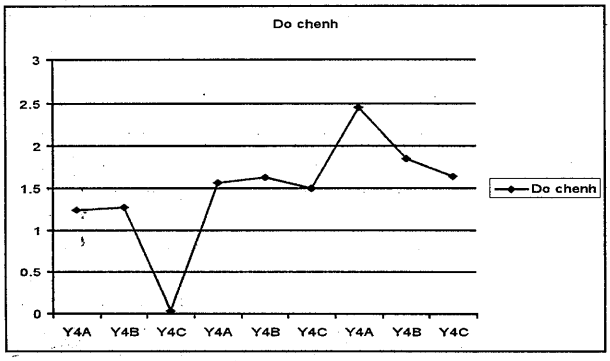


Biểu đồ 3.3. Phân bố điểm số lý thuyết Nhi của ba lớp Y4 năm học 2007-2008

3.2. Độ khó của đề thi lý thuyết nhi dành cho sinh viên Y4 qua 3 năm học từ 2005-2008:

Bảng 3.4. Độ khó của đề thi lý thuyết Nhi dành cho sinh viên Y4 qua 3 năm học từ 2005-2008 phản ánh qua độ chênh giữa TBSH và TBLT

Năm học	NH 05-06			NH 06-07			NH 07-08		
Lớp	Y4A	Y4B	Y4C	Y4A	Y4B	Y4C	Y4A	Y4B	Y4C
Độ chênh	1,25	1,28	0,03	1,56	1,63	1,50	2,46	1,85	1,64



Biểu đồ 3.4. Biến thiên của độ chênh giữa TBSH và TBLT trong điểm số thi lý thuyết Nhi của sinh viên Y4 qua 3 năm học từ 2005-2008

3.3. Độ phân biệt của đề thi lý thuyết Nhi dành cho sinh viên Y4 qua 3 năm học từ 2005-2008:

Bảng 3.5. Độ phân biệt của đề thi lý thuyết Nhi dành cho sinh viên Y4 qua 3 năm học 2005-2008 phản ánh qua sự phân bố chuẩn của điểm số

Năm học	NH 05-06			NH 06-07			NH 07-08		
Lớp	Y4A	Y4B	Y4C	Y4A	Y4B	Y4C	Y4A	Y4B	Y4C
Phân bố chuẩn	+	-	+	-	+	-	-	-	-

4. BÀN LUẬN

4. 1. Tình hình chung về điểm số môn lý thuyết Nhi của 3 lớp Y4 qua 3 năm học liên tiếp từ 2005-2008:

4.1.1. Năm học 05-06:

4.1.1.1. Lớp Y4A: Mặc dù điểm số thấp nhất là 4, cao nhất là 8, nhưng trung vị là 6 và trung bình số học là 6,25. Qua biểu đồ 3.1 cho thấy bách phân vị thứ 25 tương ứng 6 điểm và bách phân vị thứ 75 tương ứng 7 điểm. Có nghĩa là phần lớn điểm số nằm ở mức 6-7 điểm. Phân bố điểm số đạt phân bố chuẩn.

4.1.1.2. Lớp Y4B: Điểm số cao nhất là 4, cao nhất là 8, nhưng trung vị là 7 và trung bình số học là 6,28. Bách phân vị thứ 25 tương ứng với 6 điểm và bách phân vị thứ 75 tương ứng với 7 điểm. Phân bố điểm số không đạt phân bố chuẩn với hiện tượng tập trung quanh 7 điểm.

4.1.1.3. Lớp Y4C: Điểm số cao nhất là 3, cao nhất là 7, nhưng trung vị là 5 và trung bình số học là 5,03. Bách phân vị thứ 25 tương ứng với 4 điểm và bách phân vị thứ 75 tương ứng với 6 điểm. Phân bố điểm số đạt phân bố chuẩn. Đây là phân bố điểm đạt yêu cầu nhất trong 3 lớp của năm học 05-06.

4.1.2. Năm học 06-07

4.1.2.1. Lớp Y4A: Điểm số cao nhất là 4, cao nhất là 8, nhưng trung vị là 7 và trung bình số học là 6,56. Bách phân vị thứ 25 tương ứng với 6 điểm và bách phân vị thứ 75 tương ứng với 7 điểm. Phân bố điểm số không đạt phân bố chuẩn với hiện tượng tập trung quanh 7 điểm.

4.1.2.2. Lớp Y4B: Tương tự lớp Y4A nhưng điểm số đạt phân bố chuẩn.

4.1.2.3. Lớp Y4C: Điểm số cao nhất là 3, cao nhất là 8, nhưng trung vị là 7 và trung bình số học là 6,50. Bách phân vị thứ 25 tương ứng với 6 điểm và bách phân vị thứ 75 tương ứng với 7 điểm. Phân bố điểm số không đạt phân bố chuẩn với hiện tượng tập trung quanh 7 điểm. Hiện tượng tương tự lớp Y4A.

4.1.3. Năm học 07-08

4.1.3.1. Lớp Y4A: Điểm số cao nhất là 5, cao nhất là 8, nhưng trung vị là 8 và trung bình số học là 7,46. Bách phân vị thứ 25 tương ứng với 7 điểm và bách phân vị thứ 75 tương ứng với 8 điểm. Phân bố điểm số không đạt phân bố chuẩn với hiện tượng tập trung quanh 8 điểm.

4.1.3.2. Lớp Y4B: Điểm số cao nhất là 4, cao nhất là 8, nhưng trung vị là 7 và trung bình số học là 6,85. Bách phân vị thứ 25 tương ứng với 6 điểm và bách phân vị thứ 75 tương ứng với 7 điểm. Phân bố điểm số không đạt phân bố chuẩn với hiện tượng tập trung quanh 7 điểm.

4.1.3.3. Lớp Y4C: Điểm số cao nhất là 4, cao nhất là 8, nhưng trung vị là 7 và trung bình số học là 6,85. Bách phân vị thứ 25 tương ứng với 6 điểm và bách phân vị thứ 75 tương ứng với 7 điểm. Phân bố điểm số không đạt phân bố chuẩn với hiện tượng tập trung quanh 7 điểm.

Như vậy, phân bố điểm số của cả 3 lớp trong năm học 07-08 đều không đạt yêu cầu với hiện tượng phân bố tập trung vào nhóm điểm số cao.

4.2. Phân tích độ khó của đề thi phản ánh qua điểm số thi trắc nghiệm khách quan lý thuyết nhi trong 3 năm học từ 2005-2008:

Bảng 3.4 và Biểu đồ 3.4 cho thấy trong năm học 05-06 độ chênh giữa TBSH và TBLT dao động trong khoảng 0,03 – 1,28. Trong đó điểm số của lớp Y4C có độ chênh thấp nhất (0,03) có nghĩa là đạt độ khó tốt nhất. Hai lớp còn lại có độ chênh cao hơn (1,25-1,28). Trong hai năm tiếp theo, độ chênh giữa TBSH và TBLT ngày càng cao. Năm học 06-07 độ chênh dao động trong khoảng 1,5-1,63 và năm học 07-08 độ chênh tiếp tục tăng lên (1,64-2,46). Năm học 05-06 là năm Bộ môn sử dụng bộ câu hỏi mới được chỉnh sửa bổ sung nên độ khó tương đối phù hợp, thậm chí khá tốt (trường hợp của lớp Y4C, lớp được thi bộ đề mới lần đầu tiên). Qua những năm tiếp theo độ khó giảm dần một cách rõ rệt. Nguyên nhân của hiện tượng này có thể do một số yếu tố sau:

- Số lượng câu hỏi được soạn mới chưa thật nhiều.
- Mặc dù Bộ môn cố gắng bảo mật những đề đã thi bằng cách thu lại đề sau khi thi nhưng sinh viên vẫn tìm cách giấu một số trang đề nên việc bảo mật những đề đã sử dụng hầu như khó thực hiện triệt để. Các kỳ thi thường tổ chức chung 2 lớp với số lượng sinh viên đông cán bộ coi thi ít, nên việc kiểm tra đề thi khi sinh viên nộp lại không được chặt chẽ lắm.
- Các câu hỏi chưa được đánh giá độ khó cũng như độ phân biệt thông qua các phần mềm phân tích độ khó, độ phân biệt của từng câu hỏi và từng đề thi (ví dụ phần mềm QUEST của Úc). Phần lớn câu hỏi trong kho câu hỏi đang sử dụng có thể chưa đạt độ khó thích hợp. Sự phân tích độ khó và độ phân biệt của từng câu hỏi sẽ giúp loại bỏ bớt những câu hỏi có độ khó hoặc độ phân biệt quá thấp hoặc quá cao [1].

4.3. Phân tích độ phân biệt của đề thi phản ánh qua điểm số thi trắc nghiệm khách quan lý thuyết nhi trong 3 năm học từ 2005-2008:

Trong năm học 2005-2006, sự phân bố chuẩn của điểm số đã đạt được ở 2 lớp Y4A và C. Nhưng trong những năm tiếp theo, cùng với độ khó ngày càng giảm thì độ phân biệt cũng kém theo. Trong năm học 2006-2007 chỉ có đề thi của lớp Y4B có độ phân biệt đạt yêu cầu. Trong năm học 2007-2008 không có đề thi nào đạt yêu cầu về độ phân biệt.

Độ phân biệt phụ thuộc chủ yếu vào độ khó của đề thi. Độ khó càng phù hợp thì độ phân biệt càng tốt.

Phương pháp này chỉ cho phép đánh giá đề thi một cách toàn bộ (độ khó và độ phân biệt của một đề thi 100 câu hỏi) chứ không cho phép phân tích độ khó và độ phân biệt của từng câu hỏi. Khi đánh giá được độ khó và độ phân biệt của từng câu hỏi chúng ta mới có thể loại bỏ những câu hỏi qua khó hay quá dễ nhằm xây dựng một kho câu hỏi phù hợp nhất cho sinh viên. Để làm việc này, hiện nay hệ thống giáo dục Úc đã đưa ra phần mềm QUEST mà một số trường trung học, cao đẳng và đại học ở nước ta bắt đầu sử dụng. Việc sử dụng phần mềm này trên quy mô toàn trường cho phép các bộ môn có được một kho đề thi chất lượng cao. Tuy nhiên, một yếu tố khác có thể gây nhiễu cho độ khó và độ phân biệt của các câu hỏi là sinh viên có thể thuộc câu hỏi nếu kho câu hỏi của từng bộ môn không đủ lớn. Do đó, để nâng cao chất lượng của việc thi trắc nghiệm khách quan chúng ta cần đồng thời gia tăng số câu hỏi lên một mức phù hợp đi đôi với rà soát độ khó và độ phân biệt của các câu hỏi đã biên soạn bằng các kỹ thuật tiên tiến (ví dụ sử dụng phần mềm QUEST)[5].

5. KẾT LUẬN

Qua đánh giá các đề thi trắc nghiệm khách

quan loại MCQ của Bộ môn Nhi trong 3 năm học bằng cách đánh giá độ khó và độ phân biệt chúng tôi rút ra một số kết luận sau:

+ Trong năm học 05-06: Chỉ một đề thi đạt độ khó phù hợp. Hai đề thi đạt độ phân biệt phù hợp.

+ Trong năm học 06-07: Không có đề thi nào đạt độ khó phù hợp. Chỉ một đề thi đạt độ phân biệt phù hợp.

+ Trong năm học 07-08: Không đề thi nào đạt độ khó và độ phân biệt phù hợp.

Độ khó và độ phân biệt của các câu hỏi trắc nghiệm khách quan (nhiều phương án chọn) giảm dần theo thời gian sau khi được đưa ra sử dụng chứng tỏ số lượng của kho câu hỏi chưa đủ nhiều.

TÀI LIỆU THAM KHẢO

1. Trần Thị Anh Đào (2008), "Sử dụng phần mềm QUEST để phân tích câu hỏi trắc nghiệm dùng trong kiểm tra đánh giá kết quả học tập ở học sinh lớp 10", *Tuyển tập Báo cáo "Hội nghị Sinh viên Nghiên cứu Khoa học" lần thứ 6 Đại học Đà Nẵng – 2008.*
2. Khoa Sư phạm, Đại Học Quốc Gia Hà Nội (2004), "Đo lường và đánh giá thành quả học tập". *Giáo dục học đại học*. Hà Nội 2004. Tr.116-12.
3. Vũ Đình Luận (2004), "Về kiểm tra đánh giá bằng câu hỏi nhiều lựa chọn (MCQ) trong môn Di truyền ở trường CĐSP", *Tạp chí Giáo dục (88),* Tr. 36 – 37.
4. Vũ Đình Luận (2002), "Xây dựng câu hỏi nhiều lựa chọn trong môn Di truyền học ở trường CĐSP", *Tạp chí Giáo dục (55),* Tr. 34 - 35.
5. Nguyễn Bảo Hoàng Thanh (2008). "Sử dụng phần mềm QUEST để phân tích câu hỏi trắc nghiệm khách quan", *Tạp Chí Khoa Học và Công Nghệ, Đại Học Đà nẵng, Số 2(55)-2008*, Tr.119-126