

# PHÂN TÍCH ĐỀ THI TRẮC NGHIỆM KHÁCH QUAN SINH HỌC LỚP 6 BẰNG PHẦN MỀM VITESTA

ThS. TRƯƠNG THỊ THANH MAI - ThS. NGUYỄN THỊ HẢI YẾN\*

1. Trắc nghiệm khách quan (TNKQ) là một trong những công cụ quan trọng giúp đánh giá chính xác, toàn diện, khách quan kết quả học tập và trình độ của học sinh (HS). Tuy nhiên, để TNKQ trở thành công cụ đo giá trị, tin cậy, cần tuân thủ những quy định khắt khe khi thiết kế các câu hỏi dựa trên khoa học về đo lường và đánh giá. Trước đây, theo lý thuyết trắc nghiệm cổ điển, *câu hỏi trắc nghiệm (CHTN)* và *đề thi trắc nghiệm (ĐTTN)* được phân tích dựa trên các tham số đặc trưng như: độ khó, độ phân biệt (tương quan điểm nhị phân), độ giá trị và độ tin cậy. Tuy nhiên, nhược điểm của phương pháp cổ điển là kết quả phân tích phụ thuộc vào kích cỡ mẫu, không đánh giá được năng lực của HS, mức độ phù hợp giữa ĐTTN với năng lực của HS.

Từ thập niên 70 đến nay, lý thuyết trắc nghiệm hiện đại dựa trên thuyết đáp ứng câu hỏi (Item Response Theory - IRT) ra đời và phát triển nhanh chóng nhờ khắc phục được những nhược điểm của lý thuyết cổ điển. IRT được ứng dụng trong các phần mềm phân tích câu hỏi, đề thi trắc nghiệm như QUEST của Hội đồng nghiên cứu giáo dục Úc (ACER) hay VITESTA của Công ty cổ phần Khoa học và công nghệ giáo dục Việt Nam. IRT sử dụng hàm đặc trưng để tính toán mối tương quan giữa năng lực của thí sinh (TS) với các đặc trưng của CHTN thông qua các tham số: độ phân biệt IRT (a), độ khó (b) và độ đoán mò (c). Việc xử lý số liệu của mô hình tính toán theo IRT là hết sức cần thiết khi mà quan điểm đánh giá theo năng lực trở thành một xu thế tất yếu trong dạy học hiện đại.

2. Với mong muốn sử dụng IRT để đánh giá chất lượng CHTN và ĐTTN, năm học 2013-2014 chúng tôi đã tiến hành thiết kế một đề kiểm tra giữa kì gồm 40 CHTN trong chương trình *Sinh học 6* và tiến hành thực nghiệm trên 990 HS tại các trường THCS thuộc huyện Hòa Vang, TP. Đà Nẵng. Các phiếu trả lời được đưa vào máy quét OFSCAN để thu được số liệu kết quả bài làm của từng TS. Dữ liệu về bài làm của TS được đưa vào phần mềm VITESTA để định cỡ ĐTTN

theo mô hình 3 tham số, dữ liệu xuất ra bao gồm các thông tin sau: các tham số của từng CHTN theo thuyết cổ điển và IRT; đường cong đặc trưng, hàm thông tin của từng CHTN và của ĐTTN dưới dạng đồ thị; giá trị năng lực của từng TS và đường cong điểm thực của ĐTTN để thu được điểm số của từng HS theo một thang điểm mong muốn (trên 10, trên 20, hoặc trên 100...). Từ đó, cung cấp thông tin về mối tương quan giữa ĐTTN và năng lực của HS. Kết quả phân tích được thể hiện như sau:

## 1) Báo cáo chung về đề thi

BÁO CÁO VỀ ĐỀ THI

DẤP ÁN: 123224413321123344112341114443423222341  
TỔNG SỐ THI SINH: 990  
TỔNG SỐ CÂU HỎI: 40

ĐINH CỎ ĐỀ THI

SAI SỐ THAM SỐ CHO PHÉP 0.001  
SỐ VÒNG LẬP ƯỚC LƯỢNG MỖI THAM SỐ: 50  
SỐ VÒNG LẬP NEWTON (JMLE): 5  
MÔ HÌNH: 3 THAM SỐ  
ĐỘ TIN CẬY CỦA PHÉP ĐO (CỔ ĐIỂN):  
THEO PHƯƠNG PHÁP CHIA ĐỐI CHẶN LỀ ĐỀ THI 0.9925904  
ĐỘ TIN CẬY PHÂN TÍCH NĂNG LỰC (PERSON SEPARATION RELIABILITY): 0.9959111  
GIÁ TRỊ TRUNG BÌNH MỨC NĂNG LỰC: 0  
ĐỘ LỆCH CHUẨN: 1  
GIÁ TRỊ TRUNG BÌNH ĐỘ KHÓ: -0.266  
ĐỘ LỆCH CHUẨN: 1.073

Các số liệu cho thấy độ tin cậy của phép đo rất cao, trong đó độ tin cậy theo cổ điển là 0.9925 và theo IRT là 0.995; ĐTTN được đánh giá là tương đối dễ so với năng lực trung bình của TS thể hiện ở giá trị âm (- 0.266).

## 2) Các chỉ số thống kê của từng câu hỏi

Bảng 1. Các tham số của CHTN số 6 và số 12

Câu 6: Nguyên nhân không làm suy giảm sự đa dạng ở thực vật là:

- A. Khai thác bừa bãi, quá mức. B. Dân số ngày càng tăng nhanh.  
C. Do biến đổi khí hậu trên trái đất. D. Xây dựng nhiều khu bảo tồn thực vật.

Bộ que:	13			
Độ phân biệt (cổ điển):	0.45955			
Độ khó (cổ điển):	0.69806			
Độ phân biệt IRT(a):	0.66273			
Độ khó IRT(b):	-0.88620			
Độ phỏng đoán IRT(c):	0.00010			
Các phương án:	A	B	C	D*
Số TS chọn:	99	105	91	682
Tỉ lệ TS chọn PA (%):	10.13	10.75	9.31	69.81
Tương quan điểm nhị phân:	-0.25274	-0.16152	-0.15669	0.45955
Giá trị t:	-8.21089	-5.14442	-4.98662	16.21924
Giá trị p:	0.00000	0.00000	0.00000	0.00000

Câu 12: Nguyên nhân khiến mưa lũ ngày càng tăng là:

- A. Mất rừng đầu nguồn. B. Chặt phá rừng bừa bãi.  
C. Đốt rừng làm nương rẫy. D. Chặt phá rừng phòng hộ.

Bộ que:	15			
Độ phân biệt (cổ điển):	0.21472			
Độ khó (cổ điển):	0.23692			
Độ phân biệt IRT(a):	0.53305			
Độ khó IRT(b):	2.12349			
Độ phỏng đoán IRT(c):	0.09172			
Các phương án:	A*	B	C	D
Số TS chọn:	231	553	112	79
Tỉ lệ TS chọn PA (%):	23.69	56.72	11.49	8.10
Tương quan điểm nhị phân:	0.21472	0.10669	-0.18459	-0.13929
Giá trị t:	6.91031	3.37287	-5.90369	-4.42145
Giá trị p:	0.00000	0.00039	0.00000	0.00001

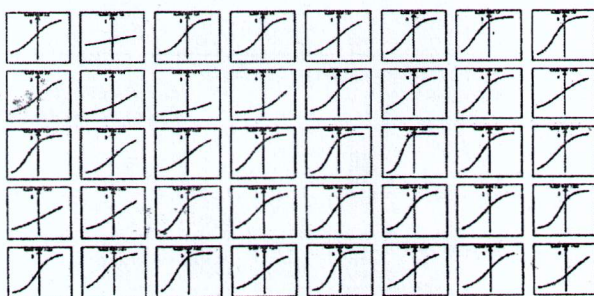
\* Trường Đại học sư phạm - Đại học Đà Nẵng



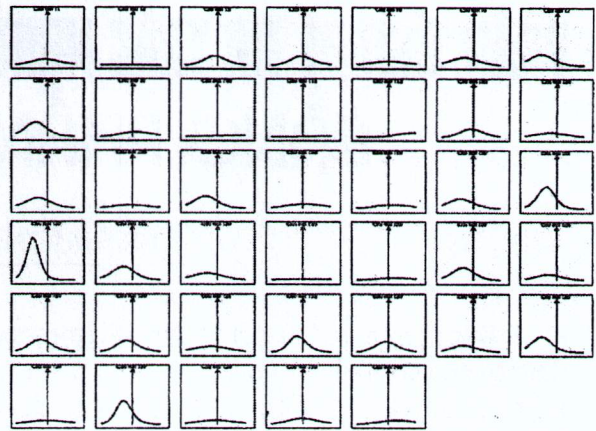
Từ các số liệu về độ khó, độ phân biệt, độ đoán mò của câu hỏi và các phương án nhiễu, xác định các CHTN có chất lượng tốt để đưa vào *ngân hàng câu hỏi* (NHCH), đồng thời, phân tích các CHTN có vấn đề để loại bỏ: - Ở câu số 6, số liệu phân tích cho thấy: độ khó cổ điển là 0,69 và độ khó IRT có giá trị âm - 0,88 chứng tỏ câu này tương đối dễ so với năng lực trung bình của TS; độ phân biệt cổ điển là 0,45 và theo IRT là 0,66 đều tốt. Các phương án nhiễu A, B, C đều đạt yêu cầu do có độ lôi cuốn đối với TS gần như nhau thể hiện ở số lượng TS chọn các phương án tương đương, có độ phân biệt (tương quan điểm nhị phân) đạt giá trị âm và giá trị p rất nhỏ ( $p < 0,05$  là tốt). Như vậy, tuy câu số 6 mặc dù khá dễ với TS nhưng các tham số thể hiện đạt yêu cầu, do đó câu này đảm bảo chất lượng, có thể sử dụng; - Câu hỏi số 12 là một trong những câu hỏi có vấn đề, thể hiện ở những điểm: độ phân biệt thấp (độ phân biệt cổ điển và IRT lần lượt là 0,21 và 0,53), câu này quá khó đối với năng lực trung bình của TS (độ khó cổ điển là 0,23, độ khó IRT là 2,12). Bên cạnh đó, TS có xu hướng đoán mò trong câu này do độ phỏng đoán IRT (c) cao (0,09). Mặt khác, số TS chọn phương án nhiễu B lại cao hơn rất nhiều so với phương án đúng A, do đó tương quan điểm nhị phân của nhiễu B có giá trị dương 0,10. Như vậy, câu này chưa tốt, cần phải loại bỏ hoặc chỉnh sửa lại.

Những CHTN chưa đạt yêu cầu có thể do các nguyên nhân: - Đáp án của CHTN là một phương án sai; - Câu đề dẫn không rõ ràng; - Các phương án nhiễu đều đúng; - Sai sót ở giai đoạn nhập số liệu kết quả bài làm. Vì thế, khi gặp các CHTN cho số liệu bất thường, giáo viên cần rà soát lại các giai đoạn nhập số liệu, ra đề thi hoặc trong quá trình dạy học có nhầm lẫn về kiến thức khiến TS chọn sai đáp án hay không. Chẳng hạn như ở câu số 12, nguyên nhân dẫn đến việc CHTN chưa đạt yêu cầu là do giáo viên đã thiết kế tất cả các phương án lựa chọn đều đúng.

### 3) Đường cong đặc trưng và hàm thông tin của từng câu hỏi



(a) Đường cong đặc trưng

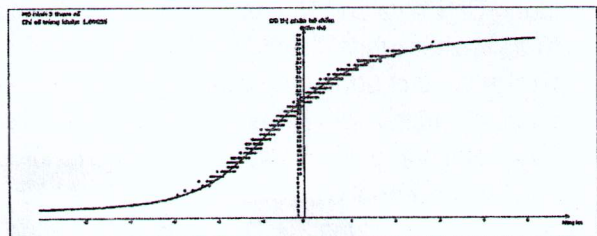


(b) Đường cong hàm thông tin

Hình 1. Ma trận đường cong đặc trưng và hàm thông tin của 40 CHTN

Các tham số của từng CHTN được biểu diễn bằng đồ thị đường cong đặc trưng và hàm thông tin (hình 1). Khi xem xét 2 loại đồ thị này, giáo viên có thể nhanh chóng xác định các câu hỏi có vấn đề bằng trực quan. Những câu hỏi bất thường sẽ thể hiện ở đường cong đặc trưng có độ dốc bé, tức có tham số a bé (độ phân biệt IRT thấp) và độ khó quá cao hoặc quá thấp (câu quá dễ hoặc quá khó), đồng thời hàm thông tin ở dạng nằm ngang chứng tỏ CHTN không thể đo được dải năng lực cần xem xét của TS. Hàm thông tin bị ảnh hưởng bởi tham số độ phân biệt và độ đoán mò của câu hỏi. Ví dụ, CHTN số 8 và 17 có cùng giá trị độ khó b là -1,1 và cùng độ đoán mò (c), tuy nhiên biên độ của hàm thông tin của câu 8 lớn hơn do có độ phân biệt nhỉnh hơn ( $a_8 = 0,88$ ,  $a_{17} = 0,81$ ). Mặt khác, tham số đoán mò (c) có giá trị càng lớn thì thông tin để đánh giá năng lực càng giảm như ở câu 17 và câu 27 có độ phân biệt xấp xỉ bằng nhau nhưng độ phỏng đoán có chênh lệch cụ thể là  $c_{17} = 0,0001$  và  $c_{27} = 0,00161$  nên biên độ hàm thông tin ở câu 17 nhỉnh hơn ở câu 27. Hình ảnh trực quan từ 2 đồ thị cho thấy trong 40 CHTN được tiến hành thực nghiệm, các câu hỏi số 2, 10, 11, 12, 16, 25, 26 là không đảm bảo chất lượng cần xem xét để điều chỉnh hoặc loại bỏ.

### 4) Phân tích sự phù hợp của ĐTTN với mô hình IRT



Hình 2. Đồ thị phân bố điểm

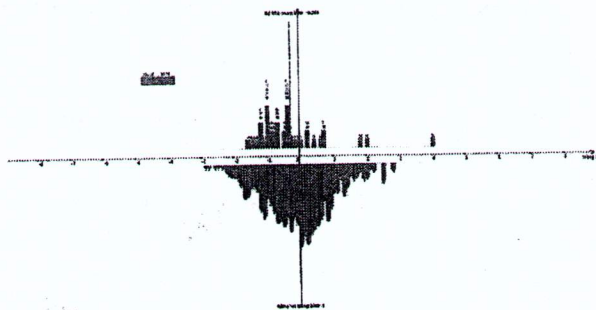


Để phân tích sự phù hợp của ĐTTN với mô hình IRT, có thể quan sát trên đồ thị đường cong điểm thực ở hình 2. Đồ thị thể hiện sự phân bố điểm thực của TS trên dải năng lực, trong đó mỗi điểm màu xanh là một hoặc nhóm các TS có cùng điểm thô và cùng mức năng lực thu được trong quá trình ước lượng. Nhìn trên đồ thị có thể thấy rằng, những TS có cùng điểm thô nhưng giá trị năng lực khác nhau. Để phân tích rõ hơn năng lực của từng TS, cần xét giá trị ước lượng năng lực của từng TS cụ thể trong bảng sau:

STT	SBD	Năng lực	Sai số chuẩn	Đúng/Tổng	Điểm quy đổi
1	160101	1.53056	0.50195	34/40	653.061
2	160102	0.59356	0.40340	29/40	559.361
3	160103	0.54102	0.39875	30/40	554.101
4	160104	1.49372	0.49762	34/40	649.371
5	160105	0.34628	0.43731	32/40	594.631
6	160106	0.54929	0.39947	30/40	554.931
7	160107	0.04596	0.36117	25/40	504.601
8	160108	-0.51037	0.23583	20/40	448.361
9	160110	-0.24316	0.34543	22/40	475.681
10	160109	0.12630	0.36643	26/40	512.631
11	160111	0.30672	0.37947	28/39	530.671
12	160112	-1.00163	0.33283	17/40	399.841
13	160113	1.19144	0.46330	33/39	619.141
14	160114	2.12234	0.57585	36/40	712.231
15	160115	2.30462	0.60014	36/40	730.461

Bảng 2. Bảng giá trị ước lượng năng lực của TS

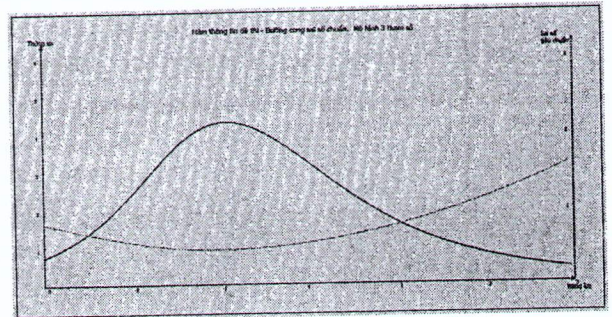
Dựa vào bảng này có thể thấy, nhiều TS có cùng số điểm thực như nhau nhưng điểm quy đổi lại khác nhau. Ví dụ, TS số 1 và 4 cùng trả lời đúng 34/40 câu nhưng năng lực khác nhau, lần lượt là 1.53 và 1.49; dẫn đến có sự sai khác về điểm quy đổi (653 và 649) do mỗi CHTN tuy có cùng điểm thô nhưng lại khác nhau về tham số độ khó và độ phân biệt IRT nên đóng góp vào kết quả ước lượng khác nhau. Điều đó chứng tỏ rằng, nếu chỉ xét điểm thô theo như lý thuyết trắc nghiệm cổ điển thì không thể đánh giá được năng lực của TS. Đây chính là một trong những ưu điểm vượt trội của mô hình trắc nghiệm IRT so với thuyết cổ điển.



Hình 3. Đồ thị phân bố độ khó và năng lực của thí sinh

Đồ thị ở hình 3 thể hiện mức độ phù hợp của ĐTTN với năng lực của TS. Năng lực và độ khó được đặt trên cùng 1 thang đo. Trên đồ thị, năng lực trung bình được quy ước bằng giá trị 0, độ khó trung bình của tất cả các CHTN (thể hiện bằng đường màu xanh đứt đoạn) có giá trị là -0.266. Như vậy, ĐTTN

tương đối dễ so với năng lực của TS. Các cột màu đỏ ở phía dưới thể hiện dải năng lực của từng TS cụ thể, các cột màu xanh ở phía trên biểu diễn độ khó của từng câu hỏi. Qua đồ thị ở hình 3 có thể thấy rằng, ĐTTN chưa phù hợp để đánh giá toàn bộ dải năng lực của TS, cụ thể là các dải năng lực từ giá trị -3 đến -1.7; từ 0.7 đến 1.8 và từ 2 đến 3 chưa có câu hỏi phù hợp để đánh giá. Đồ thị cũng thể hiện CHTN số 11 không phù hợp với mô hình do nằm ngoài dải năng lực của TS.



Hình 4. Hàm thông tin đề thi - đường cong sai số chuẩn

Đồ thị ở hình 4 thể hiện mối tương quan giữa đường cong sai số chuẩn (đường nét mảnh) và hàm thông tin đề thi (đường nét đậm). Đường cong hàm thông tin của ĐTTN trên cho thấy ĐTTN này có khả năng đo chính xác nhất ở trong khoảng năng lực trung bình và dưới mức trung bình do trong khoảng năng lực đó có sai số nhỏ nhất.

\*\*\*

Từ những số liệu trên có thể kết luận rằng, VITESTA là một trong những phần mềm có hiệu quả trong việc phân tích CHTN và ĐTTN trong mối tương quan với năng lực của TS theo lý thuyết đáp ứng câu hỏi. Phần mềm không chỉ là công cụ hữu dụng cho giáo viên khi xây dựng các ĐTTN và NHCH có chất lượng mà còn giúp giáo viên đánh giá chính xác năng lực của từng HS, từ đó có những biện pháp kịp thời để giúp đỡ HS trong quá trình học tập. □

#### Tài liệu tham khảo

1. Lâm Quang Thiệp. **Đo lường trong giáo dục, lý thuyết và ứng dụng**. NXB Đại học quốc gia, H. 2011.
2. Lâm Quang Thiệp - Lâm Ngọc Minh - Lê Mạnh Tấn - Vũ Đình Bồng. "Phần mềm Vitesta và việc phân tích số liệu trắc nghiệm". *Tạp chí Giáo dục*, số 176/2007, tr. 10-12.
3. Nguyễn Thùy Linh - Bùi Mỹ Hạnh. "Nghiên cứu một số đặc điểm câu hỏi trắc nghiệm nhiều lựa chọn

(Xem tiếp trang 50)



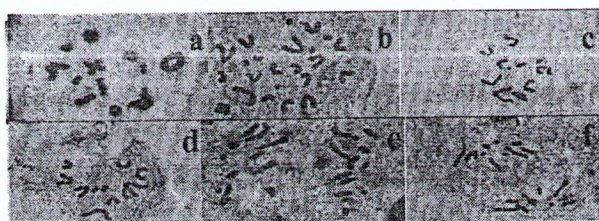
b) *Loại nước bằng cồn*. Có thể loại nước bằng cách để khô tự nhiên từ 1-2 ngày nhưng qua thực nghiệm, các tiêu bản không đồng nhất khi cố định, mẫu khô nhanh hoặc chậm tùy vào lượng nước khi lên kính. Ngoài ra, thuốc nhuộm bị oxy hóa nên màu sắc của NST bị biến đổi. Đối với cách loại nước bằng cồn, đặt lame đã được tách lamelle vào cồn có nồng độ tăng dần để loại nước khỏi TB. Thời gian ngâm mẫu trong các lọ cồn tùy theo các đối tượng khác nhau. Quá trình loại nước nếu không tốt sẽ dẫn đến trường hợp nước còn trong TB sẽ làm tiêu bản bị hư sau vài ngày, nếu loại nước quá nhanh thì TB sẽ nhanh chóng bị teo lại không quan sát được NST.

Qua khảo sát, thời gian loại nước ở các nồng độ cồn cho kết quả tốt nhất là: cồn 40° - 15 giây → 50° - 15 giây → 60° - 15 giây → 70° - 15 giây → 80° - 15 giây → 90° - 10 giây → cồn tuyệt đối - 5 giây.

Theo Võ Thị Thanh Phương (2012), cần dùng xylene hoặc n-butanol để loại nước hoặc làm sáng mẫu. Nhưng nếu sử dụng xylene không cẩn thận thì có thể gây độc cho người sử dụng. Qua thực nghiệm cho thấy có thể bỏ qua bước này mà kết quả vẫn không làm giảm màu sắc, độ tương phản và độ bền màu của tiêu bản.

c) *Dán mẫu*. Mẫu sau khi loại nước có thể dán bằng euparon hoặc baume canada pha trong xylene. Đề tài chọn canada balsam pha trong xylene với tỉ lệ 1:1 để dán tiêu bản đã cho kết quả tốt. Nhỏ 1 giọt canada balsam lên mẫu đã loại nước sau đó đẩy lamelle lại và để khô tự nhiên sau vài ngày.

3) *Khảo sát độ bền màu*. Quan sát tiêu bản cố định được thực hiện vào tháng 03/2012 đến 02/2014 vẫn có độ bền màu tốt.



Hình 3. Hình ảnh NST ở châu chấu (*Oxya chinensis*) (X1.000)

a. Kì đầu I b. Kì giữa I c-d. Kì giữa II e-f. Kì sau II

\* \* \*

Đề tài đã đề xuất được quy trình thực hiện tiêu bản hiển vi tạm thời và cố định, trong đó có thể quan sát rõ hình thái và số lượng NST. Nồng độ nhược trương thích hợp nhất là natri citrate 0,4% trong 30 phút, giúp NST phân tán đều trong TB. Nhuộm NST bằng aceto-

orcein 2% trong 15 phút ở 60°C hoặc 30 phút ở nhiệt độ phòng cho kết quả tốt. Để tiêu bản trong ngăn đá tủ lạnh từ 1-2 giờ sẽ giúp tách rời lamelle khỏi lame dễ dàng mà không làm xê dịch mẫu. Tiêu bản cố định vẫn đạt yêu cầu khi lược bỏ công đoạn loại nước bằng xylene hoặc n-butanol. Dán mẫu bằng canada balsam pha với xylene theo tỉ lệ 1:1. □

#### Tài liệu tham khảo

1. Huỳnh Thị Ngọc Nhân - Kiều Ngọc Ân - Mai Thị Tuyết. **Thực tập di truyền cơ sở**. NXB Đại học quốc gia TP. Hồ Chí Minh, 2004.
2. Lê Đình Trung - Đặng Hữu Lan. **Di truyền học**. NXB Giáo dục, H. 2000.
3. Nguyễn Nghĩa Thìn. **Thực vật có hoa**. NXB Đại học quốc gia, H. 2006.
4. Trần Tú Ngà. **Giáo trình thực tập di truyền và chọn giống**. NXB Nông nghiệp, H. 1982.
5. Võ Thị Thanh Phương. “Khảo sát số lượng nhiễm sắc ở tế bào thực vật và tế bào động vật bằng phương pháp xử lí sốc nhược trương”. Tạp chí *Khoa học Đại học Cần Thơ*, 2012, số 21; tr. 198-208.
6. Vũ Đức Lưu - Nguyễn Minh Công. **Giáo trình Di truyền học**. NXB Đại học sư phạm, H. 2007.

#### SUMMARY

Based on the procedures performed microscope slide in grasshopper meiosis, the author improved a number of actions to create the template serves morphological observation and counting the number of chromosomes of the male grasshopper. It can be used as a reference for teachers and students to contribute additional sources of fixed specimens to enhance the teaching quality of genetics exercises in schools.

## Phân tích đề thi trắc nghiệm...

(Tiếp theo trang 47)

bằng lí thuyết đáp ứng câu hỏi”. Tạp chí *Nghiên cứu y học*, số 68/2010, tr.130-137.

4. Nguyễn Bảo Hoàng Thanh. “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 2/2008, tr. 119-126.

#### SUMMARY

This paper presents the results of the analysis of a Sixth Grade Biology objective test with the support of VITESTA software. This software evaluating the quality of questions and objective tests in the interaction with the student's competency based on Item Response Theory.