

Avaliação do Sistema de Avaliação de Alunos da Disciplina de Patologia II – 1ª Parte

LEONOR LEVY*

Resumo

Neste trabalho são apresentados os resultados obtidos pelos alunos do 1º Curso de Enfermagem da Universidade Atlântica no primeiro tempo de avaliação. Os métodos de avaliação utilizados neste primeiro tempo são um exame teórico constituído por um teste e a avaliação de um relatório sobre o trabalho de campo efectuado.

Constata-se que os resultados obtidos pelos alunos no exame teórico revelam um enviesamento à direita, enquanto os resultados obtidos através da avaliação do relatório do trabalho de campo e a nota total da Disciplina de Patologia II, revelam um enviesamento à esquerda, não havendo correlação entre os resultados obtidos no teste e no trabalho de campo. Constatou-se ainda que a maior parte das perguntas de escolha múltipla e resposta única tinham poder discriminativo e um baixo Índice de Facilidade.

Palavras-Chave: Patologia II, Avaliação, Índice de Discriminação, Índice de Facilidade.

Summary

Evaluation of the Assessment System of Pathology II Students – Part 1

In this part, the results obtained by students in the different types of assessment methods in the first time are presented. The theoretical examination is biased towards the right and the results of the work field and the final marks are biased to the left.

The majority of multiple choice questions are discriminatory and Have a low Facility index. There is no correlation between the results on the theoretical exam and the work field.

Key-Words: Pathology II, Assessment, Evaluation, Discrimination Index, Facility Index.

Introdução

A Disciplina de Patologia II, está integrada no 2º Ano do Curso de Enfermagem da Universidade Atlântica (UATLA), situada em Barcarena. Trata-se de uma Universidade privada com diversos cursos, tendo iniciado o 1º Curso de Enfermagem em 2001.

A autora deste trabalho foi convidada pelo Reitor da UATLA para ser a Coordenadora e Regente da Disciplina de Patologia II daquela Universidade, tendo tido liberdade para escolher os Docentes e organizar o Programa de Ensino-Prendizagem da Disciplina em questão. Os Docentes da Disciplina de Patologia II foram escolhidos pelas suas qualidades científicas e pedagógicas. Todos os Docentes desta Disciplina são profissionais altamente qualificados e trabalham no Hospital de Santa Maria e/ou na Faculdade de Medicina de Lisboa.

No primeiro trabalho desta série apresentou-se o Programa de Ensino-Aprendizagem da Disciplina de Patologia II, do 1º Curso de Enfermagem da Universidade Atlântica (UATLA), a sua filosofia, sistema e metodologia de ensino, conteúdos e objectivos, assim como o sistema de avaliação⁽¹⁾.

A segunda parte do estudo é constituída pela apresentação dos resultados obtidos no 1º tempo de avaliação dos alunos.

Material e Métodos

O trabalho de campo deste estudo foi efectuado no 1º semestre de 2002.

No fim do Programa de Ensino-Aprendizagem os alunos foram avaliados através de um exame teórico e através de um relatório sobre o trabalho de campo.

O exame teórico para avaliação das aquisições no

domínio cognitivo, consistiu na realização de um teste escrito de vinte e nove perguntas de escolha múltipla e resposta única e ainda de três perguntas de *certo e errado*.

Foram avaliados o Índice de Discriminação e o Índice de Facilidade para cada uma das perguntas de escolha múltipla e resposta única do teste.

O relatório de um trabalho de campo individual teve como objectivo a descrição das ocorrências observadas durante um turno de enfermagem em diferentes hospitais, em recém-nascidos patológicos.

Crítérios de aprovação:

- Comparência nas aulas teórico-práticas (80% das aulas)
- Obtenção de uma classificação mínima de 8,5 valores no teste (máximo de 17 valores).
- Obtenção de uma classificação mínima do relatório do trabalho de campo individual de 1,5 valores (máximo de 3 valores)
- O critério mínimo de aprovação, ou seja a soma das duas formas de avaliação foi de dez valores.

A classificação final (máximo de 20 valores) foi determinada pela ponderação;

- Exame teórico 17 valores (máximo)
- Relatório do trabalho de campo 3 valores (máximo)

O estudo estatístico foi efectuado através do SPSS11,5 para Windows.

O estudo estatístico iniciou-se com a análise univariada e a apresentação das variáveis intervaladas em termos de estatística descritiva e representações gráficas^(2,4).

As variáveis de tipo intervalado foram estudadas em termos de estatística descritiva e foram ainda apresentadas em histograma com curva normal sobreposta, em diagrama de *caule e folhas e caixa de bigodes*^(2,4).

As associações de variáveis intervaladas foram estudadas através de regressão linear, tendo sido escolhido como limiar de significância, o valor de p de 0,05^(2,4).

Resultados

Exame teórico

O exame teórico consistiu num teste de vinte e nove perguntas de escolha múltipla e de resposta única e três perguntas de resposta curta.

A maior nota possível do teste foi de dezassete valores. Compareceram ao exame cinquenta e quatro alunos.

Nota total do teste

A variável nota total do teste não segue uma curva normal, sendo leptocúrtica e com uma distribuição moderadamente assimétrica negativa.

A média, mediana e trimédia a 5% não têm valores não muito diferentes; no entanto, a mediana é superior à média, sugerindo um enviesamento à direita (Quadros I e II).

A análise da simetria, através do quociente enviesamento e o seu erro padrão é igual a -1,144, não permite, no entanto rejeitar a simetria da curva.

Quadro I

		Estatística	Erro padrão	
nota total do teste	Média	8,49016	,226551	
	Intervalo com 95% de confiança para a média	Limite inferior	8,03576	
		Limite superior	8,94457	
	Trimédia a 5%	8,52623		
	Mediana	8,76563		
	Variância	2,772		
	Desvio padrão	1,664806		
	Mínimo	4,250		
	Máximo	12,484		
	Dispersão	8,234		
	Dispersão interquartis	2,12500		
	Simetria	-,372	,325	
Curtose	,405	,639		

Quadro II

Percentil 25	7,43
Percentil 50	8,76
Percentil 75	9,56

O histograma com a sobreposição da curva normal também sugere que a curva tem uma assimetria negativa com um enviesamento à direita (Gráfico 1).

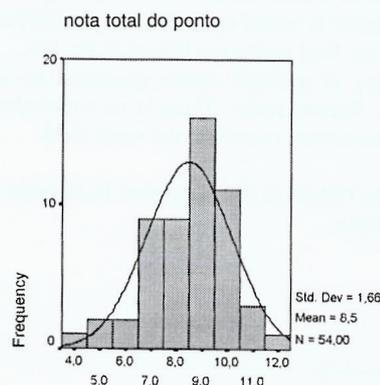


Gráfico 1

O diagrama de caule e folhas também sugere uma assimetria negativa ou enviesamento à direita (Gráfico 2).

nota total do teste	Caule e Folhas	
Frequência		
1,00	Extremos	(=<4,3)
2,00	4 .	77
1,00	5 .	8
6,00	6 .	166699
11,00	7 .	14447777799
7,00	8 .	2257777
17,00	9 .	00000000222555588
6,00	10 .	003336
2,00	11 .	14
1,00	12 .	4
Stem width:	1,000	
Each leaf:	1 case (s)	

Gráfico 2

A caixa de bigodes mostra um *outlier* ou valor extremo, mostrando que a mediana está colocada ligeiramente acima do meio da caixa, sugerindo também um enviesamento à direita (Gráfico 3)

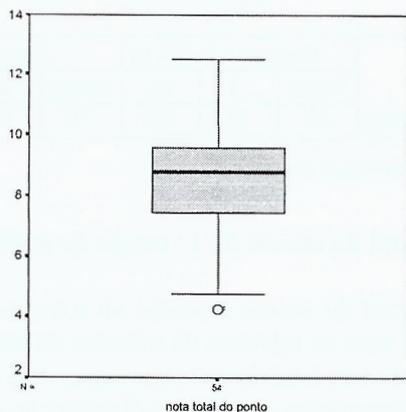


Gráfico 3

O teste de Kolmogorov-Smirnov (K-S) mostra que a curva não é normal (Quadro III).

Quadro III

	Kolmogorov-Smirnov (a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
nota total do ponto	,121	54	,046	,978	54	,426

a Lilliefors Significance Correction

Avaliou-se também o Índice de Discriminação das perguntas de escolha múltipla e resposta única do teste.

Pode verificar-se que 18 das 29 perguntas tiveram o poder de discriminar os estudantes (Quadro IV)⁽⁵⁾.

Quadro IV

Índice de Discriminação	
P1 = 0,44	P16 = 0,66
P2 = 0,44	P17 = 0,66
P3 = 0,14	P18 = 0,44
P4 = 0,14	P19 = 0,37
P5 = 0,44	P20 = 0,0
P6 = 0,66	P21 = 0,74
P7 = 0,14	P22 = 0,37
P8 = 0,66	P23 = 0,44
P9 = 0,37	P24 = 0,07
P10 = 0,29	P25 = 0,37
P11 = 0,37	P26 = 0,29
P12 = 0,07	P27 = - 0,07
P13 = 0,14	P28 = 0,51
P14 = 0,59	P29 = 0,44
P15 = 0,22	

Também foi determinado o índice de Facilidade (5), revelando que 75% dos alunos responderam acertadamente a sete das vinte e nove perguntas de escolha múltipla e resposta única (Quadro V).

Quadro V

Índice de Facilidade	
P1 = 85,1	P16 = 87,0
P2 = 64,8	P17 = 44,4
P3 = 9,3	P18 = 33,3
P4 = 7,4	P19 = 37,0
P5 = 44,5	P20 = 25,9
P6 = 66,7	P21 = 51,9
P7 = 83,3	P22 = 27,8
P8 = 81,5	P23 = 53,7
P9 = 96,3	P24 = 25,9
P10 = 94,4	P25 = 74,1
P11 = 61,1	P26 = 70,4
P12 = 13,0	P27 = 22,2
P13 = 16,7	P28 = 35,2
P14 = 87,0	P29 = 35,2
P15 = 48,1	

Trabalho de campo

O trabalho de campo consistiu na observação das ocorrências existentes durante um turno de enfermagem, em recém-nascidos patológicos em dois estabelecimentos de saúde, o Hospital de Santa Maria e a Maternidade Dr. Alfredo da Costa.

A nota máxima possível do trabalho de campo foi de três valores.

A variável *nota do trabalho de campo* não segue uma curva normal, sendo leptocúrtica e com uma distribuição moderadamente assimétrica positiva.

A média, mediana e trimédia a 5% não têm valores não muito diferentes; no entanto, a mediana é inferior à média, sugerindo um enviesamento à esquerda (Quadro VI).

A análise da simetria, através do quociente enviesa-

mento e o seu erro padrão é igual a 1,477, não permitindo, no entanto, rejeitar a simetria da curva.

Quadro VI

		Statistic	Std. Error	
nota trabalho de campo	Média	2,6033	,03088	
	Intervalo com 95% de confiança para a média	Limite inferior	2,5402	
		Limite superior	2,6665	
	Trimédia a 5%	2,6000		
	Mediana	2,5000		
	Variância	,029		
	Desvio padrão	,16914		
	Mínimo	2,30		
	Máximo	2,90		
	Dispersão	,60		
	Dispersão interquartis	,3000		
	Simetria	,631	,427	
	Curtose	-1,039	,833	

Quadro VII

Percentil 25	2,50
Percentil 50	2,50
Percentil 75	2,80

O histograma com a curva normal sobreposta também sugere uma assimetria positiva com enviesamento à direita (Gráfico 4), assim como o diagrama de caule e folhas (Gráfico 5).

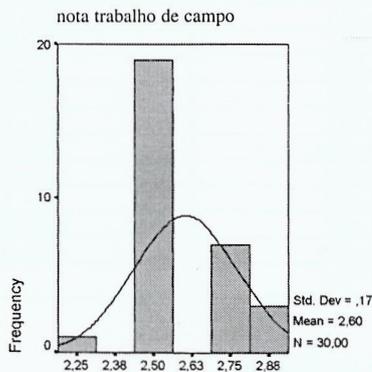


Gráfico 4

Frequency	Stem & Leaf
1,00	23 . 0
,00	24 .
19,00	25 . 00000000000000000000
,00	26 .
,00	27 .
7,00	28 . 0000000
3,00	29 . 000

Stem width: ,10
Each leaf: 1 case (s)

Gráfico 5

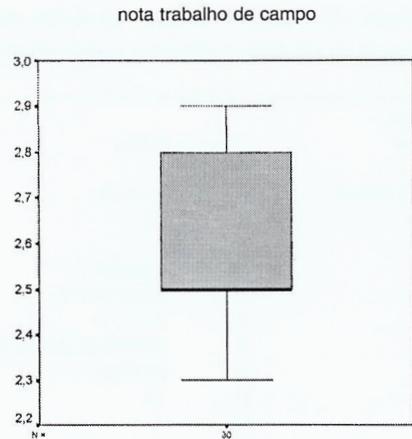


Gráfico 6

A caixa de bigodes mostra que a mediana é igual ao percentil 25 (Gráfico 6)

A análise da normalidade através do teste de Kolmogorov-Smirnov (K-S) mostra que a curva não é normal (Quadro VII).

Quadro VII

nota trabalho de campo	Kolmogorov-Smirnov (a)			Shapiro Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
nota trabalho de campo	,396	30	,000	,730	30	,000

a Lilliefors Significance Correction

Nota total do exame do 1º tempo de avaliação

A nota total do exame consistiu no somatório da nota do teste e da nota do relatório do trabalho de campo.

A variável nota total do teste não segue uma curva normal, sendo leptocúrtica e com uma distribuição moderadamente assimétrica positiva.

Quadro VIII

		Statistic	Std. Error	
nota total - soma teste 1º tempo e trabalho de campo	Média	12,1897	,19846	
	Intervalo com 95% de confiança para a média	Limite inferior	11,7838	
		Limite superior	12,5956	
	Trimédia a 5%	12,1080		
	Mediana	11,7900		
	Variância	1,182		
	Desvio padrão	1,08701		
	Mínimo	11,00		
	Máximo	15,25		
	Dispersão	4,25		
	Dispersão interquartis	1,4450		
	Simetria	1,167	,427	
	Curtose	,707	,833	

A média, mediana e trimédia a 5% não têm valores não muito diferentes; no entanto, a mediana é inferior à média, sugerindo um enviesamento à esquerda (Quadro VIII).

A análise da simetria, através do quociente enviesamento e o seu erro padrão é igual a 2,7, permitindo rejeitar a simetria da curva.

Quadro IX

Percentil 25	11,47
Percentil 50	11,79
Percentil 75	12,91

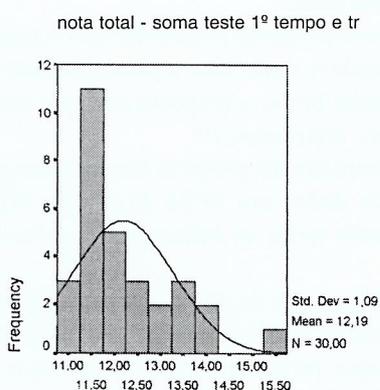


Gráfico 7

O histograma com a sobreposição da curva normal sugere que a curva tem uma assimetria positiva com um enviesamento à esquerda (Gráfico 7)

O diagrama de caule e folhas também sugere uma assimetria positiva ou enviesamento à esquerda (Gráfico 8)

nota total - soma teste 1º tempo e trabalho de campo Stem-and-Leaf Plot

Frequency	Stem & Leaf
7,00	11 . 0012223
10,00	11 . 555555778
3,00	12 . 003
3,00	12 . 568
2,00	13 . 12
2,00	13 . 66
2,00	14 . 00
1,00 Extremes	(>=15, 3)

Stem width: 1,00
Each leaf: 1 case (s)

Gráfico 8

A caixa de bigodes mostra um outlier ou valor extremo, mostrando que a mediana está colocada bastante

abaixo do meio da caixa, perto do percentil 25, sugerindo também um enviesamento à esquerda (Gráfico 9)

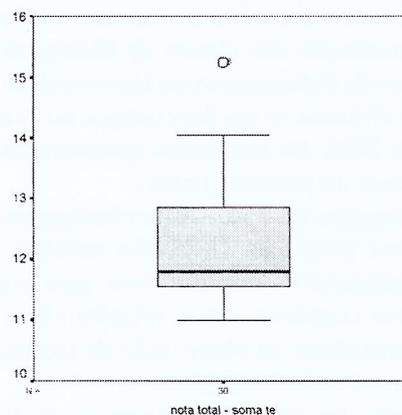


Gráfico 9

A avaliação da normalidade da curva através do teste de Kolmogorov-Smirnov (K-S), mostra que a curva não é normal (Quadro X)

Quadro X

	Kolmogorov-Smirnov (a)			Shapiro Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
nota total - soma teste 1º tempo e, trabalho de campo	,207	30	,002	,867	30	,001

a Lilliefors Significance Correction

Correlações entre os dois métodos de avaliação dos alunos

Fez-se a correlação entre a nota do teste e a nota do trabalho de campo, pontuando cada uma destas variáveis sobre vinte valores, verificando-se que não existe uma correlação entre os dois valores obtidos, nota total do primeiro teste e nota do trabalho de campo (Quadro XI).

Quadro XI

		nota total primeiro teste	nota trabalho de campo
Pearson Correlation	nota total primeiro teste	1,000	,184
	nota trabalho de campo	,184	1,000
Sig (2 tailed)	nota total primeiro teste		,332
	nota trabalho de campo	,332	

Discussão

Tratou-se de um estudo prospectivo, exploratório e não experimental com o objectivo de fazermos a avaliação dos sistemas de avaliação dos alunos da Disciplina Patologia II, do 1º Curso de enfermagem da Universidade Atlântica.

O estudo efectuou-se em dois tempos ou "timings", em Novembro de 2002. Os resultados apresentados neste trabalho referem-se ao primeiro tempo.

Os instrumentos de observação consistiram num teste de vinte e nove perguntas de escolha múltipla e resposta única e três perguntas de resposta curta, para avaliação dos alunos no nível cognitivo e num relatório de um trabalho de campo, consistindo na observação de um recém-nascido patológico em meio hospitalar.

A avaliação dos alunos da Disciplina de Patologia II foi baseada em critérios e não na norma.

A avaliação baseada no critério (CRT), baseia-se na determinação de objectivos e no grau da sua concretização. Com este processo, os estudantes não são comparados com uns com os outros, mas consigo próprios, o CRT compara o desempenho do estudante com um critério e não com os resultados da classe. Uma crítica que se faz ao CRT é a de que o CRT não encoraja a excelência, mas apenas garante que a maior parte dos alunos atinge o nível mínimo de aceitação, avaliando o conhecimento no seu nível mais baixo e não em níveis mais elevados como o da interpretação e resolução de problemas^(6,8).

Foram definidos dois critérios mínimos de aprovação: como critério mínimo dos testes foi definida a nota de 8,5 valores, correspondendo a 50% da nota máxima possível e como critério mínimo do relatório do trabalho de campo foi definida a nota de 1,5 valores, correspondendo a 50% da máxima nota possível; como critério mínimo de aprovação na Disciplina de Patologia II foi definida a nota de dez valores, correspondendo a 50% da nota máxima possível.

Para cada pergunta de escolha múltipla e resposta única do teste foram observados os índices de Discriminação e de Facilidade.

O Índice de Discriminação é a medida do poder que tem cada pergunta do teste para discriminar entre os mais capazes (ou os que têm mais conhecimentos) e os menos capazes. Um poder discriminativo menor que 0,2, faz com que as perguntas sejam retiradas e entre 0,2 e 0,3, sugere que as perguntas sejam revistas⁽⁵⁾.

O Índice de Discriminação varia entre +1 e -1; um Índice de Discriminação negativo significa que a pergunta está mal feita. Para um *item* ser encarado como um bom discriminador, o seu poder de discriminação deverá ser positivo e igual ou superior a 0,3.

Pode haver um enviesamento possível no que respeita ao Índice de Discriminação, se uma dada pergunta for formulada de uma maneira ambígua ou mal construída⁽⁹⁾.

O Índice de Facilidade define a proporção dos candidatos que responderam de maneira acertada a uma dada pergunta. O Índice de Facilidade varia de 0 a 1, podendo ser apresentado em percentagens.

Um índice de Facilidade igual a 1 significa que todos os candidatos acertaram na pergunta ou seja que a pergunta, era muito fácil; um índice de Facilidade igual a 0 significa que nenhum candidato acertou na pergunta.

Um alto Índice de Facilidade (pergunta muito fácil) ou baixo Índice de Facilidade (pergunta muito difícil) depende de vários factores, incluindo a maneira como a pergunta é formulada. Numa pergunta de escolha múltipla e resposta única, o Índice de Facilidade é muito influenciado pelo *distractors*.

Uma mesma pergunta poderá ter como resultado diferentes desempenhos, consoante a plausibilidade dos *distractors* e a maneira como a resposta correcta está escondida no meio desses *distractors*⁽⁵⁾.

A complexidade do projecto levou à criação de diferentes bases de dados em SPSS 11,5 para Windows, que pudessem conter todas as informações colhidas ao longo do estudo.

Para a elaboração do texto foi especialmente importante o Manual de Pedro Serrano⁽¹¹⁾, enquanto que para o estudo estatístico, foram utilizados os Manuais de Anthony Walsh⁽²⁾, de Pestana e Gageiro⁽³⁾ e de Reis e Melo⁽⁴⁾.

O tratamento estatístico foi da responsabilidade da autora do estudo.

Para este efeito, a autora do projecto frequentou vários cursos de introdução à investigação e cursos de estatística.

A análise estatística iniciou-se com a apresentação das variáveis intervaladas em termos de estatística descritiva^(2,4).

As variáveis de tipo intervalado foram estudadas em termos de estatística descritiva e foram ainda apresentadas em histograma com curva normal sobreposta, em diagrama de *caule e folhas e caixa de bigodes*⁽¹¹⁾.

As associações de variáveis intervaladas foram estudadas através de regressão, tendo sido escolhido como limiar de significância, o valor de p de 0,05.

A análise da apresentação em termos de estatística descritiva e das representações gráficas da variável *nota total do teste* sugere um ligeiro enviesamento à direita.

Ao analisarmos as notas, no entanto, verifica-se que a média é de 8,490, e a variância de 2,77, correspondendo assim, na maior parte dos casos, a notas baixas.

Ao observarmos o Índice de Discriminação, verificamos que a maior parte das perguntas tiveram o poder de discriminar entre os melhores e os piores alunos.

O Índice de Facilidade do teste também foi observado, verificando-se que a maior parte das perguntas tinham um baixo Índice de Facilidade, ou sejam, correspondiam a perguntas difíceis.

Podemos especular que estas notas e estes Índices

podem ter várias explicações: os Docentes deste curso são profissionais muito competentes e muito exigentes, habituados a construir perguntas para alunos do Curso de Medicina da Faculdade de Medicina de Lisboa, sendo, assim, as suas perguntas difíceis para alunos de um curso de Enfermagem de uma Universidade privada; os próprios alunos poderão ter um nível inferior aos de alunos de outras escolas, ou poderão não ter investido muito no estudo desta Disciplina.

A nota do relatório de trabalho de campo, mostra um enviesamento à esquerda; no entanto, ao analisarmos as notas verifica-se a maior parte dos alunos tem uma nota igual ou superior a 2,5 valores, quando a nota máxima possível é de 3 valores.

As condições em que decorreu o trabalho de campo pode explicar estes valores; num dos hospitais onde decorreu o trabalho de campo, houve dificuldades na sua concretização, passando o trabalho de campo a ter a cotação máxima de apenas 3 valores; todos os relatórios foram avaliados pela mesma pessoa, o que diminui os enviesamentos; no entanto, dado que os relatórios só podiam ter duas páginas, torna-se mais difícil a discriminação entre os alunos.

A nota total do exame, correspondeu ao somatório da nota do teste e da nota do relatório do trabalho de campo, mostrando um enviesamento à esquerda, podendo significar ou um baixo nível dos alunos, ou uma exigência grande dos Docentes da Disciplina de Patologia II.

Fez-se ainda a comparação entre os dois métodos de avaliação dos alunos, através de uma regressão linear, verificando-se que não existe uma correlação entre os dois métodos de avaliação.

A falta de correlação existente entre os dois tipos de avaliação é preocupante; afinal, apesar dos dois tipos de avaliação avaliarem dimensões diferentes do desempenho dos alunos, para o exercício da profissão de enfermeiro é essencial a integração harmoniosa das diferentes dimensões, para alcançar a excelência no exercício da Arte e Ciência da Enfermagem.

Bibliografia

1. Levy L. apresentação UATLA
2. Walsh A. Statistics for the Social Sciences. With Computer Applications. New York Harper & Row, Publishers, 1988.
3. Pestana MH, Gageiro JN. Análise de Dados para Ciências Sociais. A complementaridade do SPSS. *Edições Silabo*, 1998.
4. Reis E, Melo P, Andrade R, Calapez T. Estatística Aplicada. Vol. 2. *Edições Silabo* 1997.
5. Sutton, R.A. (1997). An Introduction to Assessment & Evaluation Processes and Procedures. *University College Cardiff*.
6. Tumbull JM. What is Normative versus Criterion-referenced Assessment. *Medical Teacher* 1989; 2: 145-150.
7. Lowry S. Assessment of students. *BMJ* 1993; 306: 51-54.
8. Lowry, S. Making change happen. *BMJ* 1993; 306: 320-322.
9. Hobsley M. Counting apples with oranges: a limitation of the discrimination index. *Medical Education* 1999; 33: 192-196.
10. Serrano P
11. Levy L. A Iliteracia em Estatística Tem Remédio. *Acta Ped Port*, 2000; 31: 401-408