

AValiação DE ATIVIDADES PRÁTICAS DE ALUNOS DE MEDICINA. MÉTODO REPETITIVO-COMPARATIVO

Rubens de Alencar¹

RESUMO

O autor procurou verificar a validade de um novo método de avaliação de atividades práticas para estudantes de medicina, sistema que poderíamos chamar de repetitivo-comparativo. Ficou demonstrado que:

1. Os alunos discordaram de si mesmos, quanto a um mesmo ponto, com apenas uma semana de diferença de tempo, entre um e outro laudo;
2. Aparentemente, ao analisarem um mesmo material, sem aviso, pela 2.^a vez, melhoram de rendimento;
3. O Professor, de forma indireta, também é automaticamente avaliado.

INTRODUÇÃO

Um dos mais sérios problemas com que se depara o professor universitário é o da avaliação da competência de seus alunos. Isto se deve a muitos fatores e a abordagem deve ser diferente, sempre que necessária.

Anualmente, ao longo de quase vinte anos de ensino, recebemos alunos para avaliação. Ao fim da mesma, freqüentemente ficamos na dúvida quanto ao aferimento, visto que seguimos roteiro predeterminado, cuja validade, a nosso ver, é sempre muito discutível.

No trabalho que desenvolvemos procuramos mostrar ao aluno seu grau de concordância com o professor, sua qualificação no ponto aferido. Simultaneamente tentamos verificar sua concordância com os colegas, revelando sua situação dentro do grupo. Por último, ao estudarmos a concordância do aluno consigo mesmo, revelamos sua coerência.

Este tipo de trabalho não é original^{1, 3, 6}, mas tem sido principalmente desenvolvido em profissionais graduados. Desconhecemos, em nosso meio, trabalho entre estudantes.

I – METODOLOGIA E CASUÍSTICA

O trabalho foi realizado no Hospital Universitário Clementino Fraga Filho (UFRJ), em 1986, durando aproximadamente um mês.

Trabalhamos com três internos acostumados a ver fundo de olho de diabéticos. A cada um deles solicitávamos que, após olhar diapositivo de retinopatia diabética, dissesse em que grau ela poderia ser classificada. Olhavam, no máximo, cinco diapositivos por dia, até atingirem o total de vinte. Após este número, sem serem avisados, passavam a examinar as mesmas fotos, para novos laudos. Deste modo verificávamos o grau de concordância do interno consigo mesmo. O laudo de cada foto, dado por um interno, era comparado com os dos outros internos e com o do oftalmologista. Assim agindo observávamos o grau de concordância dos internos entre si e com o oftalmologista (laudo-padrão).

A retinopatia diabética, classificada por Balantyne em quatro grupos, foi reduzida, para simplificação, em apenas dois. Assim:

Grau I (nosso) = Graus I e II de Balantyne

Grau II (nosso) = Graus III e IV de Balantyne

Os internos só precisavam dizer se a retinopatia era grau I ou II.

Para análise das fotos todos dispunham do mesmo tempo, do mesmo visor e viram exatamente os mesmos diapositivos. A análise dos resultados foi feita de acordo com Sackett⁷ e Fleis⁴.

Os internos foram chamados de A, E e F.

¹ Professor Adjunto de Clínica Médica, Faculdade de Medicina, UFRJ.

II RESULTADOS

Tabela 1 – Laudos dos internos e do professor

Foto n.º	A (1ª)	A (2ª)	E (1ª)	E (2ª)	F (1ª)	F (2ª)	Professor
20	2	2	1	1	1	1	1
24	2	1	1	1	1	1	1
25	2	2	2	2	2	1	1
26	2	2	1	1	2	2	2
28	1	2	2	2	2	2	2
29	2	2	2	2	2	2	2*
30	1	1	1	1	1	1	1*
31	1	1	1	1	1	1	1*
33	1	2	2	2	2	2	2
35	2	1	1	1	1	1	2
36	1	2	2	1	2	2	2
38	2	2	1	1	2	2	2
39	2	2	1	2	2	1	1
40	2	2	2	2	2	2	2*
41	1	2	1	1	1	2	1
42	2	2	1	2	2	2	2
47	2	2	2	2	1	1	2
51	2	2	1	2	2	2	2
53	2	2	1	1	1	1	1
55	2	2	1	1	2	2	2

* = 100% de concordância

Tabela 2

Concordância (K) dos internos com o Professor

Prof.	A (1ª) A (2ª)		E (1ª) E (2ª)		F (1ª) F (2ª)	
	0,13	0,32	0,36	0,32	0,58	0,72

Tabela 3

Concordância intrapessoal (internos) na 1ª tentativa

1ª tentativa	A E	1ª tentativa	
		E	F
		0,16	0,16 0,34

Tabela 4

Concordância intrapessoal dos internos, na 2ª tentativa

2ª tentativa	A E	2ª tentativa	
		E	F
		0,47	0,47 0,49

Tabela 5

Concordância intrapessoal dos internos, nas duas tentativas

1ª tentativa	A E F	2ª tentativa		
		A	E	F
		0,16	0,52	0,69

III – DISCUSSÃO

A verificação da concordância entre as diversas medidas foi feita pelo índice k^4 , descrito por Cohen². Segundo Landis e Koch⁵, para a maioria dos estudos clínicos, valores de k superiores a 0,75 são considerados excelentes; entre 0,40 e 0,75, aceitáveis; inferiores a 0,40 não são merecedores de crédito. Convém lembrar que o índice k é uma medida de correlação que procura excluir o fator acaso.

A tabela 1 nos mostra que houve muita variação, da primeira para a segunda tentativa. Dos vinte diapositivos examinados somente houve concordância completa, entre professor e alunos, nas duas tentativas, em apenas quatro deles (Nos. 29, 30, 31 e 40), ou seja: 20%.

A análise da tabela 2 mostra a baixa concordância dos alunos com o professor. Verifica-se, também, que os internos A e F apresentaram nítida melhora, da 1ª para a 2ª tentativa. O interno E apresentou diminuição da concordância, mas de modo não significativo.

A tabela 3 revela a concordância entre internos, na 1ª tentativa. Todos tiveram k inferior a 0,40. De acordo com Landis & Koch⁵, estes valores poderiam resultar apenas do acaso. Já na 2ª tentativa (tabela 4), esses índices melhoraram porém mantiveram-se em torno de 0,47 — 0,49. À primeira vista esta melhora deveu-se a maior interesse pelo assunto fazendo com que, na 2ª tentativa, acertassem mais e ficassem mais coerentes. Pelo pequeno número de dados achamos necessária uma repetição do estudo, com maior número de participantes, e procurar fazer com que os alunos estudassem entre uma e outra tentativa.

Por último, para verificar a coerência nos laudos, fizemos com que todos os slides fossem revistos, nas mesmas circunstâncias. As concordâncias variam de 0,16 a 0,69.

Todos estes resultados, no fim do trabalho, foram mostrados aos internos e ao professor. Os primeiros, porque nunca pensaram que estivessem tão mal e fossem tão incoerentes; o professor, porque julgava ter sido claro em suas exposições prévias e ficou de discutir o problema com seus colegas, visando à revisão de metodologia.

SUMMARY

The author tried a new method of evaluating competence of medical students, giving them a simple

and objective material to analyse. After a week the same students received the same material and made a new analysis. Both studies were compared with a previously made analysis made by a Professor. The results showed that the students were very incoherent, but probably able to improve in just a week and became convinced that they were still needing a lot of practice. The Professor who taught that matter confessed to be very surprised and decided to revise his methodology.

REFERÊNCIAS BIBLIOGRÁFICAS

1. AOKI, N. et alii. Observer variability and reproducibility for fubdosopic findings. *Japan Circ. J.*, 41:11, 1979.
2. COHEN, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, 20:37, 1960.
3. DETRE, K.M.; WRIGHT, E. & MURPHY, M.L. Observer agreement in evaluating coronary angiograms.
4. FLEISS, J.L. The measurement of interrater agreement. In: ——. *Statistical methods for rates and proportions*. 2. ed. New York, Wiley — Interscience, s.d.
5. LANDIS, J.R. & KOCH, G.G. The measurement of observer agreement for categorical data. *Biometrika*, 33:159, 1977.
6. MEAD, T.W.; GARDNER, M.J. & ROSCH, J. Observer variability in recording the peripheral pulses. *Br. Heart J.*, 30:661, 1968.
7. SACKETT, D.L. et alii. Clinical disagreement: I — How often it occurs and why. *Can. Med. Assoc. J.*, 20:

ENDEREÇO DO AUTOR:

Faculdade de Medicina UFRJ
Edifício do Centro de Ciências da Saúde
Bloco K — 2º andar
Ilha da Cidade Universitária
21910 — Rio de Janeiro — RJ