

Padronização da Correção de Questões Dissertativas para Professores de Saúde Coletiva do Curso de Medicina em uma Instituição de Ensino Superior do Oeste do Paraná

▮ Alana Schirmer Caporal *

▮ Maria Rosa Machado Prado **

▮ Ivair Rogério Bini ***

▮ Christian Boller ****

Resumo

Este trabalho refere-se ao uso de questões dissertativas como método de avaliação, e como sua correção pode variar entre diferentes avaliadores, resultando em discrepâncias de conceitos. Para evitar tal problemática, é possível utilizar ferramentas de padronização de correção, como a rubrica (do inglês *rubric*). Assim, esse estudo objetivou averiguar se, com o uso da rubrica, a correção se torna mais clara e homogênea. Para tanto, selecionou-se dez professores de saúde coletiva de uma IES do Oeste do Paraná para que realizassem a correção de uma prova dissertativa aplicada a 33 acadêmicos de medicina. Metade desses professores realizou a correção pelo método tradicional, enquanto os demais corrigiram utilizando a rubrica. Após análise estatística de comparação, evidenciou-se que a variabilidade das médias pela correção tradicional foi maior. Nas primeiras questões as notas são mais díspares e nas últimas a discrepância diminui, indicando que, com o método tradicional, os professores tendem a alterar o rigor no final da correção. Dessa forma, as rubricas podem tornar o processo de correção mais preciso e transparente, favorecendo o *feedback* entre professores e acadêmicos.

Palavras-chave: Rubrica. Rubricas de pontuação. Prova dissertativa.

* Médica e Mestranda no Programa de Pós-Graduação em Ensino nas Ciências da Saúde, Faculdades Pequeno Príncipe. Docente Faculdade Assis Gurgacz; E-mail: alanaschirmer@hotmail.com.

** Farmacêutica e Doutora em Biotecnologia, Universidade Federal do Paraná. Docente do PPG em Ensino nas Ciências da Saúde da Faculdades Pequeno Príncipe; E-mail: mrosaprado@hotmail.com.

*** Matemático, Especialista em Estatística. Docente nos cursos de graduação da Faculdades Pequeno Príncipe; E-mail: rogeriobini@gmail.com

**** Farmacêutico e Doutor em Biotecnologia aplicada à Saúde da Criança e do Adolescente na Faculdades Pequeno Príncipe. Docente do PPG em Ensino nas Ciências da Saúde da Faculdades Pequeno Príncipe e discente em Licenciatura em Filosofia da FAE – Centro Universitário; E-mail: christian.boller@gmail.com.

1. Introdução

Diversos modelos de avaliação estão disponíveis para o docente avaliar o progresso do estudante. Entre eles, a aplicação de avaliações com questões discursivas permite verificar, além do conhecimento teórico, a estrutura linguística e o poder de persuasão dos argumentos utilizados pelo discente. Estes últimos são componentes subjetivos do processo avaliativo e podem variar de acordo com o docente, tempo de curso, cansaço físico, entre outros, o que pode resultar em discrepâncias no conceito (nota) fornecido ao discente por diferentes docentes, e também em diferentes graus de exigência entre a primeira e a última prova corrigida por um mesmo docente (MENDES, 2012).

Levando em consideração a necessidade de tornar a correção de questões dissertativas mais coerentes, objetivas e com menor discrepância estatística entre diferentes avaliadores, faz-se necessária a utilização de ferramentas que padronizem a correção. Dentre essas, a ferramenta a rubrica apresenta-se como um esquema de pontuação descritiva, desenvolvida por professores, ou outros avaliadores, para orientar a análise dos produtos ou processos de esforços dos alunos (BROOKHART, 1999).

A rubrica pode ser utilizada como instrumento formativo e somativo do processo de ensino-aprendizagem, facilitando a avaliação em áreas consideradas subjetivas, complexas e imprecisas. O progresso do estudante é verificado mediante critérios que qualificam gradualmente o aprendizado, conhecimentos e competências esperadas para um discente em determinado nível de ensino (GATICA-LARA; URIBARREN-BERRUETA, 2013).

Considerando a importância da avaliação de questões discursivas na formação do estudante, o presente trabalho pretendeu verificar se o uso das rubricas na correção de provas dissertativas reduz a discrepância dos escores inferidos aos alunos por diferentes avaliadores.

2. O processo avaliativo no ensino superior

O processo de aprendizagem pode ser definido como o modo com que os seres humanos adquirem novos conhecimentos, desenvolvem competências e refletem sobre seu comportamento. A aprendizagem ocorre por diferentes meios, entre eles pelas

técnicas de ensino, cujo aperfeiçoamento é tema de estudos constantes (ZEFERINO; PASSERI, 2007).

Os desafios que a educação enfrenta atualmente se referem a três aprendizagens básicas: aprender a conhecer, aprender a fazer, aprender a ser. Neste sentido, a aprendizagem efetiva envolve três domínios básicos, cognitivo, psicomotor e afetivo, que, de forma integrada, podem ser denominados de competências na aprendizagem. Essas competências são as habilidades que o estudante adquire por meio da assimilação do conteúdo, da aplicação prática deste e da atitude adotada frente ao conhecimento (ZEFERINO; PASSERI, 2007).

Vale ressaltar qual saber que o estudante adquiriu competência para a profissão ao final da disciplina, módulo ou curso é laborioso. Embora não seja prática unânime, a avaliação da aprendizagem não é mais compreendida como o ato de medir o desempenho de um estudante. Avaliar a aprendizagem significa determinar a aquisição de competências em seus aspectos quantitativos e qualitativos (STEIN et al., 2005).

Nesse sentido, a avaliação é complexa porque envolve modos de agir culturalmente estabelecidos, implicando incômodos inevitáveis como comparar, classificar, selecionar o objeto (ou ser humano) em apreciação. Minimizar esse incômodo significa inverter a lógica, orientando a avaliação educativa no sentido de garantir a aprendizagem do aluno (BORBA; FERRI; HOSTINS, 2007).

No ensino superior, a motivação do acadêmico para refazer seu trajeto de aprendizagem depende do seu conhecimento a respeito dos objetivos a serem alcançados, do roteiro que precisa ser percorrido para isso e dos critérios utilizados para avaliar seu desempenho nesse percurso. Esse esforço é influenciado diretamente pela postura do avaliador à medida que identifica, organiza e divulga os objetivos de aprendizagem e os critérios de avaliação e revela abertura para redimensionar a sua ação de acordo com o desempenho dos acadêmicos no decorrer do processo (BORBA; FERRI; HOSTINS, 2007).

3. Formas de Avaliação

As diferentes formas de avaliação são elementos centrais do processo de ensino-aprendizagem de qualquer programa educacional, e devem ser bem planejadas e

implementadas em todas as propostas curriculares. Uma avaliação adequada do rendimento acadêmico guarda estreita relação com a competência e a capacitação do profissional que será entregue à sociedade. Desta forma, destacam-se três formas de avaliação principais, diagnóstica, somativa e formativa (OLIVEIRA, 2002).

A avaliação diagnóstica tem como propósito constatar se os alunos possuem os conhecimentos básicos e imprescindíveis às novas aprendizagens. Essas ideias vão ao encontro do que postula Vygotsky (1984), que reforça que as atividades na sala de aula são influenciadas pela sociedade, ao mesmo tempo que a influenciam.

Já a avaliação formativa é realizada no decorrer do curso com o objetivo de verificar se os alunos dominam gradativamente cada etapa proposta. Neste intuito, é importante considerar que os alunos progredirão se compreenderem suas possibilidades e fragilidades, e se souberem como se relacionar com elas. Esta forma de avaliação visa informar o professor e o aluno sobre o rendimento da aprendizagem e a localização das deficiências na organização do ensino para possibilitar correção e recuperação, bem como representa o principal meio pelo qual o estudante passa a conhecer seus erros e acertos (OLIVEIRA, 2002).

A última função é representada pela avaliação somativa, cujo objetivo é determinar o grau de domínio do aluno em uma área de aprendizagem. Isto permite outorgar uma qualificação que pode ser utilizada como um sinal de credibilidade da aprendizagem realizada. Também tem o propósito de classificar os alunos ao final de um período de aprendizagem, semestre, ano, mês ou curso, de acordo com os níveis de aproveitamento (BLOOM; HASTINGS; MADAU, 1975).

Numa estrutura curricular formulada por objetivos e pré-requisitos, dificilmente esta forma de avaliação não classifica o estudante. Embora muito se fale que a avaliação somativa é punitiva, parece não haver um consenso em como desmistificar esse termo, pois, na realidade, ela precisa ser seletiva para não comprometer o desenvolvimento de novas aprendizagens no estudante (PACHECO, 1998).

4. Subjetividade da Avaliação

As provas subjetivas são aquelas em que se exige dos candidatos conhecimentos sobre um determinado tema proposto, cujo critério de correção leva em consideração não somente o conteúdo das respostas apresentadas, mas também a argumentação

lógica, a coerência, a adequação vocabular, bem como a correção gramatical e a ortografia. Alguns desses critérios de avaliação dependem do avaliador e da significância que ele atribui a cada item, ou a resposta como um todo (MONTEIRO, 2014).

O avaliador, enquanto ser humano tem suas preferências, gostos e conhecimentos, que podem interferir na correção e nos conceitos que atribui a cada resposta, e por isto está suscetível a falhas. Há problemas na avaliação que se devem ao perfil do avaliador, à sua confiabilidade (consistência na avaliação de textos de mesma qualidade), à sua tolerância ao cansaço, à sua interação com outros avaliadores e a seus erros na avaliação (MENDES, 2012).

A avaliação de questões discursivas é um processo cansativo e psicologicamente desgastante, cujo grau de tolerância varia entre indivíduos. Segundo Wolcott (1998), em situação de cansaço, a atenção do avaliador começa a reduzir, levando-o a pontuar ou a penalizar um texto, sacrificando a precisão pela rapidez.

Estudos de Erickson (2001), Charney (1984) e Stach (1987) mostram que as pontuações feitas por vários avaliadores não coincidem no que diz respeito a erros gramaticais, vocabulário, complexidade sintática ou mesmo aparência da resposta; já o estudo de Huot (1993) aponta para o fato de que avaliadores não treinados se apoiam mais em aspectos estruturais do que em aspectos comunicativos.

Além dos problemas relatados, há também o erro dos avaliadores, que podem ocorrer devido a fatores afetivos e cognitivos (WOLLCOTT, 1998). Os erros de halo ocorrem quando o avaliador lê uma questão de qualidade média depois de ter lido uma, ou mais de uma, de excelente qualidade, tendendo a penalizar mais do que necessário devido à resposta mediana seguinte ou vice-versa; erros por leniência aparecem quando o avaliador tende a ser muito clemente ou muito rigoroso; já os erros por restrição de calibragem existem quando o avaliador tende a evitar notas muito altas ou muito baixas. Ademais, há de se levar em consideração que o mesmo avaliador, no mesmo contexto de correção, no mesmo dia, pode atribuir notas muito diferentes às mesmas provas (MENDES, 2012).

Para evitar a influência de fatores subjetivos durante uma avaliação, esta pode ser segmentada em categorias de critérios, que são examinadas individualmente. Esse tipo de avaliação é útil quando há várias dimensões a serem avaliadas num trabalho, como,

por exemplo, as questões dissertativas. Para que a segmentação funcione, deve-se padronizar os critérios de aplicação e de correção das questões propostas, a fim de melhor comparar as respostas apresentadas pelos candidatos, bem como para fazer valer os princípios da igualdade, da impessoalidade, da moralidade, da eficiência, da motivação, da razoabilidade e da proporcionalidade (MONTEIRO, 2014).

5. Padronização de correção de avaliações dissertativas

O termo rubricas tem sua origem na palavra inglesa “*Rules*” (regras) que são estabelecidas no início do processo e orientam os alunos por quais caminhos podem/devem trilhar para potencializar sua aprendizagem (FELICIO; CARITÁ; OLIVEIRA NETO, 2013).

A rubrica é uma forma de pontuação que enumera os critérios para a realização de uma determinada atividade solicitada pelo docente e como será avaliada. Além disso, ajuda os estudantes a identificarem como sua atividade será avaliada e especifica o nível de desempenho esperado com vista aos vários níveis de qualidade, a partir dos quais os estudantes reconhecem se conseguiram atingir os objetivos esperados em suas próprias atividades antes de entregá-las (CRUZ; NUNES, 2009).

As rubricas esclarecem o que é importante a ser avaliado e contêm descrições qualitativas de critérios de desempenho, os quais funcionam bem dentro do processo de avaliação formativa, além de fornecerem um feedback sobre a eficácia da participação de um estudante em uma determinada atividade (JÁCOME, 2013).

O uso de rubricas possibilita que o professor desenvolva e sistematize critérios e indicadores de avaliação, buscando, assim, reduzir a subjetividade do processo. São ferramentas que servem para auxiliar o professor a construir critérios avaliativos mais transparentes e coerentes em relação a objetivos de aprendizagem determinados. Funcionam, portanto, como um instrumento de avaliação formativa, permitindo ainda o envolvimento dos alunos no processo de aprendizagem e de avaliação (RAPOSO; MARTÍNEZ, 2011).

Dois componentes são essenciais para o desenvolvimento de uma rubrica: os critérios de desempenho e os pesos atribuídos a cada critério. Os critérios de desempenho identificam elementos específicos ou dimensões da atividade que será avaliada pela

rubrica e fornecem diretrizes, regras ou princípios pelos quais os estudantes serão julgados. Os pesos descrevem as diferenças qualitativas dos estudantes, avaliando o desempenho, como, por exemplo, péssimo, ruim, regular, bom ou excelente (FELICIO; CARITÁ; OLIVEIRA NETO, 2013).

De forma geral, uma rubrica é formada por duas dimensões: a primeira é geralmente denominada de dimensão de qualidade, que podem ser temas, etapas ou o que quer que se pretenda avaliar; a segunda contempla qualificadores pré-determinados em uma escala de pontuação (FELICIO; CARITÁ; OLIVEIRA NETO, 2013).

O uso das rubricas facilita a avaliação de trabalhos complexos, torna a correção mais objetiva e transparente, além de possuir categorização, devido à quantidade de níveis adequada para a avaliação. A rubrica possibilita padronização - uma vez que elenca, alguns critérios, os quais são utilizados para todas as provas - e propicia clareza, já que clarifica as expectativas se for utilizada como um meio de comunicação com os alunos (BIAGIOTTI, 2005).

As rubricas, se bem utilizadas, diminuem a subjetividade envolvida na correção de qualquer componente dissertativo e, dessa forma, permitem uma avaliação precisa e justa, promove a compreensão e traz transparência na comunicação entre alunos e professores, resultando melhoria no processo ensino/aprendizagem (JÁCOME, 2013).

6. Metodologia

A pesquisa foi desenvolvida com docentes do módulo de Programa de Aprendizagem Saúde e Sociedade (PASS), também chamada de saúde coletiva, do curso de Medicina de uma Faculdade Privada, localizada na região oeste do Estado do Paraná. O trabalho foi submetido e aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos e aprovado sob número CAAE 58363016.6.0000.5580. A população em estudo compreendeu dez professores da disciplina, que assinaram o Termo de Consentimento Livre e Esclarecido para realizar a correção das provas discursivas aplicadas aos acadêmicos do segundo período de medicina. Do total de professores, cinco realizaram correção convencional e os outros utilizaram a rubrica, elaborada pela pesquisadora.

Aplicou-se, então, uma prova dissertativa, contendo 10 questões, para uma turma de saúde coletiva do 2º período do referido curso, perfazendo um total de 33 estudantes. Os

professores que participaram do estudo, ao final da correção, atribuíram uma nota à questão, variando de zero a 100. Ao final das correções foi realizado um comparativo entre as notas e observado se houve diferença estatística entre as notas atribuídas pelos professores que corrigiram as questões pela forma tradicional e pelo uso das rubricas.

Para a correção com as rubricas, utilizou-se uma tabela com quatro critérios de avaliação, contendo: 1) conhecimento teórico, 2) persuasão do argumento, 3) coerência/adequação vocabular e 4) gramática/ortografia. Assim, para cada um desses itens o professor poderia pontuar a qualidade em uma escala de zero a três. Ao final, foram somadas as pontuações atribuídas para concluir a nota.

Quanto ao item - conhecimento teórico/conteúdo – o professor poderia atribuir três pontos caso considerasse a resposta como “a resposta é clara e correta. A ideia principal se destaca e é apoiada por informações detalhadas”; atribuir dois pontos se considerasse “a ideia principal é clara, mas a informação de apoio é geral”; um ponto se “a ideia principal é um pouco clara, mas há uma necessidade de mais informações para apoio”; ou zero se “a ideia principal não é clara, existe um conjunto de informação aparentemente aleatória”. A pontuação máxima para esse quesito tinha uma equivalência de 30 pontos, sendo que três pontos equivalem a 30 em nota, dois pontos equivalem a 20 em nota, um ponto equivale a 10 e zero não pontuava.

Em relação ao item - persuasão do argumento/argumentação lógica – o corretor poderia atribuir três pontos se considerasse que “todos os argumentos são convincentes e embasados em literatura atual. Detalhes são colocados em uma ordem lógica”; atribuir dois pontos se “a maioria dos argumentos são convincentes e embasados em literatura atual. Detalhes quase sempre colocados em ordem lógica”; um ponto se “os argumentos são claros, mas não são embasados em literatura atual. Detalhes distraem o leitor”; ou zero se “os argumentos não são claros e convincentes. Há pouco sentido na organização da escrita”. Esse quesito também conta com a pontuação máxima equivalente a 30.

O próximo item analisado era - coerência/adequação vocabular - em que o professor também poderia atribuir pontos de zero a três, sendo que a resposta que conquistava três pontos era aquela em que “todas as frases são bem construídas com estrutura e vocabulário variado”; dois pontos se “a maioria das sentenças são bem construídas com

estrutura e vocabulário variado”; um ponto se “a maioria das sentenças são bem construídas, mas têm uma estrutura semelhante e pobre em vocabulário”; ou zero se “as sentenças são mal estruturadas, falta vocabulário, parece incompleto”. A pontuação máxima em equivalência era de 20 pontos na nota, dessa forma, se o professor pontuasse três equivaleria a uma nota 20, se pontuasse dois equivaleria a 10, um ponto equivaleria a 5 e zero não pontuava.

O último item analisado era a - gramática/ortografia – podendo-se atribuir três pontos se o “escritor não faz erros de gramática ou ortografia que distraem o leitor a partir do conteúdo”; dois pontos se o “escritor faz 1-2 erros de gramática ou ortografia que distraem o leitor a partir do conteúdo”; um ponto se o “escritor faz 3-4 erros de gramática ou ortografia que distraem o leitor a partir do conteúdo”; ou zero se o “escritor faz mais de 4 erros de gramática ou ortografia que distraem o leitor a partir do conteúdo”. Esse quesito também possui como nota máxima 20 pontos em nota, conforme explicado no item anterior.

Os dados foram analisados no software *GraphPad Prism* versão 5.0. Para cada estudante foram inferidas dez notas divididas em dois grupos, cinco para professores que corrigiram a prova no modelo tradicional e cinco com o uso das rubricas. Para cada grupo foi obtida a média e o desvio padrão com o qual foi realizado o teste estatístico. A metodologia de análise adotada foi o teste T de *Student* para amostras pareadas, que permite avaliar dois grupos distintos de valores obtidos por sistemas diferentes. Este método permite verificar se existe diferença estatística a um nível de significância de 95%. A partir dos dados obtidos entre a correção tradicional e por meio das rubricas, foi determinado o índice kappa, que verifica a discrepância entre as notas atribuídas aos estudantes por meio das diferentes formas de correção.

7. Resultados e discussão

As notas médias dos acadêmicos obtidas pelas duas formas de correção se encontram resumidas na Tabela 1.

Tabela 1 - Dados comparativos entre as notas atribuídas para a correção tradicional e por meio da técnica *rubric*

		Média	
		Tradicional (T)	Rubric (R)
1	Acadêmico 1	63,8	58,8
2	Acadêmico 2	59,2	54,3
3	Acadêmico 3	84,0	76,7
4	Acadêmico 4	83,3	80,4
5	Acadêmico 5	64,7	63,5
6	Acadêmico 6	76,7	65,3
7	Acadêmico 7	79,2	77,1
8	Acadêmico 8	74,2	75,5
9	Acadêmico 9	83,8	78,5
10	Acadêmico 10	67,0	78,5
11	Acadêmico 11	88,3	64,8
12	Acadêmico 12	78,0	84,9
13	Acadêmico 13	72,4	76,3
14	Acadêmico 14	81,4	78,5
15	Acadêmico 15	72,5	78,0
16	Acadêmico 16	49,8	74,5
17	Acadêmico 17	61,4	52,3
18	Acadêmico 18	64,0	63,4
19	Acadêmico 19	72,1	72,6
20	Acadêmico 20	78,6	72,5
21	Acadêmico 21	58,2	78,4
22	Acadêmico 22	79,5	53,9
23	Acadêmico 23	59,3	77,1
24	Acadêmico 24	79,0	57,2
25	Acadêmico 25	82,5	74,4
26	Acadêmico 26	76,4	77,1
27	Acadêmico 27	81,1	77,8
28	Acadêmico 28	78,9	76,7
29	Acadêmico 29	56,5	77,7
30	Acadêmico 30	60,8	56,4
31	Acadêmico 31	63,7	57,2
32	Acadêmico 32	72,4	57,1
33	Acadêmico 33	66,3	63,4
Média		71,8	70,0

Fonte: Os autores (2017).

Pode-se observar que a correção com o uso da rubrica é diferente da correção tradicional. A correção com a rubrica diminuiu, em média, dois pontos na nota, na escala de

zero a 100, em relação à correção pela forma tradicional, com intervalo de confiança entre 0,71 - 3,57, sendo esta diferença estatisticamente significativa ($p < 0,05$).

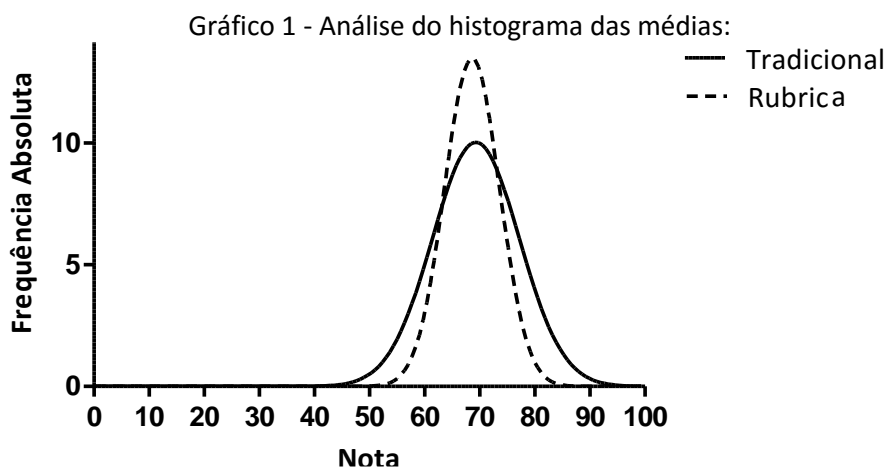
Um exemplo de discrepância na nota foi em relação ao acadêmico 11. Ao ser analisada sua prova observou-se algumas diferenças em relação aos dois tipos de correção. Em uma das questões em que era solicitado conceituar intersectorialidade, um professor utilizando a forma tradicional pontuou 80, enquanto outro, utilizando a rubrica, pontuou 50 (sendo zero em relação ao conhecimento teórico, 20 em relação à persuasão do argumento, 10 sobre a adequação vocabular e 20 quanto à ortografia). Fica claro neste contexto, que a correção pela rubrica foi mais exigente, sem, no entanto, afastar-se dos critérios de correção adotados. Por outro a técnica tradicional apresentou nota alta que não pôde ser justificada.

No caso do acadêmico 16, em uma de suas questões, dois professores que corrigiram pelo método tradicional atribuíram zero à sua resposta. Porém, nenhum dos professores que corrigiu utilizando a ferramenta *rubric* pontuou menos que 35. Isso poderia ser justificado, pois, como a nota com a rubrica é atribuída levando-se em consideração quatro conceitos previamente estabelecidos, sendo dois deles relacionados ao vocabulário e à gramática/ortografia, mesmo uma resposta completamente equivocada em relação ao conhecimento exigido para a questão poderia pontuar caso o vocabulário estivesse adequado, não houvesse erros gramaticais e tivesse boa ortografia.

Dessa forma, mesmo a resposta do acadêmico tendo fugido completamente do tema proposto, ainda assim, houve pontuação devido aos critérios da rubrica não estarem relacionadas ao conhecimento teórico sobre o assunto. Por esse motivo, no momento da construção de uma rubrica, deve-se saber que o objetivo é projetar um sistema de qualificação e interpretação consistente com as expectativas da disciplina, e isso requer um estudo profundo e, em alguns casos, um suporte especializado (MENDES, 2012).

Segundo Ferreira Sobrinho (1997), em relação às desvantagens do uso de questões dissertativas que devem ser consideradas está na possibilidade de poderem não ser representativas do rendimento do aluno e valorizarem mais a habilidade de redação do que o conteúdo. No caso do acadêmico 16, de fato houve uma supervalorização da habilidade de redação, tendo em vista que o acadêmico não pontuou no quesito “conhecimento teórico”.

As médias das notas pela correção tradicional variaram entre 49,8 a 88,3, enquanto as médias das notas pela correção com as rubricas variaram entre 52,3 a 84,9. Com isso, podemos observar que a variabilidade das médias, utilizando a rubrica, foi menor.



Fonte: Os autores (2017).

A análise do Gráfico 1 permite avaliar que a correção utilizando a rubrica possui maior uniformidade, pois um maior número de estudantes (frequência absoluta) apresentou notas semelhantes, o que pode demonstrar uniformidade na avaliação. Já na correção pelo sistema tradicional o desvio foi maior, o que pode ser observado por uma base mais larga, indicando que as notas dos estudantes variaram mais um em relação ao outro.

A correlação entre as notas utilizando a rubrica ocorre da seguinte maneira: nas primeiras questões corrigidas as notas são mais discrepantes, com o decorrer da correção, nas últimas questões, as notas começam a ficar mais semelhantes entre as duas técnicas. Como a rubrica possui critério e segue o padrão de notas, é um indicativo de que, com a correção tradicional, os professores tendem a alterar o rigor no final da correção.

A partir dos dados obtidos entre a correção tradicional e por meio da rubrica, foi possível determinar o Índice *Kappa*, que verifica a discrepância entre as notas atribuídas aos estudantes por meio das diferentes formas de correção. Esta medida de concordância tem como valor máximo o 1, em que o valor 1 representa total concordância e os valores próximos e até abaixo de 0 indicam nenhuma concordância, ou a concordância foi exatamente a esperada pelo acaso. Um eventual valor de *Kappa* menor que zero sugere que a concordância encontrada foi menor do que aquela esperada por acaso (LANDIS; KOCH, 1977). O quadro 1 demonstra a aplicação da correlação na questão 1 da prova.

Quadro 1 - Índice *Kappa* para a questão 1

Q1		Tradicional											Soma
		0	1	2	3	4	5	6	7	8	9	10	
Rubrica	0			1									1
	1		1		1	1							3
	2												0
	3						1						1
	4						2	1					3
	5						1	3					4
	6								2	1			3
	7										1		1
	8									3	5	1	9
	9										4	4	8
	10												0
Soma		0	1	1	1	1	4	4	2	4	10	5	33

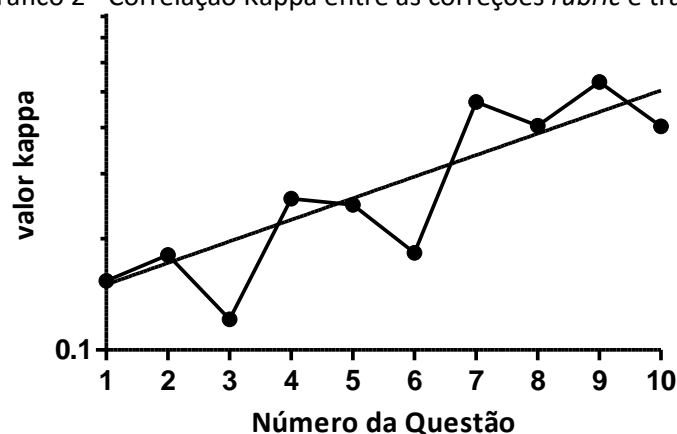
Fonte: Os autores (2017).

A análise do Índice *Kappa* demonstra que, quando dois valores são iguais em situações distintas (tradicional x rubrica), o valor de *Kappa* tende a 1, indicando a proximidade das notas com as duas técnicas. Por outro lado, quanto mais distante da região central, é indicativo de diferença estatística entre as duas técnicas.

Landis e Koch (1977), sugerem a seguinte interpretação quanto à qualidade da classificação da estatística *Kappa*: valores <0,00 - qualidade da classificação péssima; 0,00-0,20 - qualidade da classificação ruim; 0,21-0,40 - qualidade da classificação razoável; 0,41-0,60 - qualidade da classificação boa; 0,61-0,80 - qualidade da classificação muito boa; e 0,81-1,00 - qualidade da classificação excelente. Essa avaliação de concordância por meio do *Kappa* é utilizada quando as escalas são categóricas e sempre quando estamos comparando dois ou mais parâmetros.

O Índice *Kappa* foi aplicado para todas as questões da prova e o Gráfico 2 apresenta os valores de correlação *Kappa* para cada questão. É possível observar que à medida que os corretores avançam na correção da prova, e o valor de *Kappa* aproxima-se de 1, se reduzem a discrepância entre as correções. A partir dos dados, foi obtido a regressão linear com equação da reta igual a $y = 0,0584x - 0,8813$, indicando que o valor de *Kappa* (y) é crescente e tende a 1 quanto maior o número de questões (x) corrigidas.

Gráfico 2 - Correlação Kappa entre as correções *rubric* e tradicional



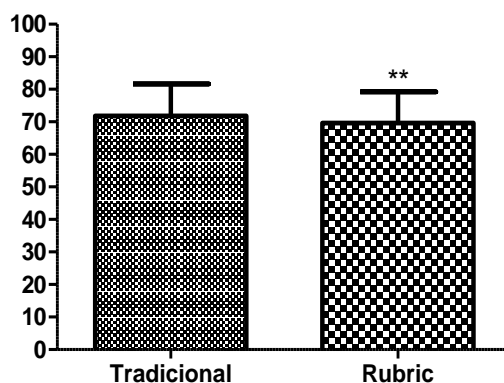
Nota: A reta representa a regressão linear dos pontos onde $y = 0,0584x - 0,8813$.

Fonte: Os autores (2017).

Observa-se, no Gráfico 2, que as primeiras questões são as que possuem pior correlação. Como as rubricas tem critério e mantém um padrão de notas, é um indicativo de que o rigor da avaliação tradicional é alterado com o decorrer da correção, fazendo com que as notas obtidas por meio dos dois métodos se aproximem nas últimas questões.

Estes dados confirmam o apontado por Mendes (2012), que postula que a tolerância ao cansaço é um fator determinante na qualidade da correção, e que quanto mais longa a avaliação, maior a probabilidade de o professor reduzir sua atenção na correção. Estes dados permitem inferir que o uso da rubrica tende a manter constante a forma de correção, pois utiliza regras definidas.

Gráfico 3 - Análise das médias dos estudantes



** Representa $p < 0,05$

Fonte: Os autores (2017).

A análise do Gráfico 3 permite observar que a nota média dos estudantes pela técnica tradicional foi ligeiramente maior que a técnica *rubric*. A correção com a ferramenta *rubric*

diminuiu, em média, dois pontos na nota, na escala de zero a 100, em relação à correção pela forma tradicional. No entanto, quando aplicado o teste T, observou-se que o desvio padrão da rubrica é menor e estatisticamente significante ($p < 0,05$). Isto aponta para a hipótese de que utilizando a rubrica as avaliações tendem a ser mais transparentes, equitativas e justas, conforme postula Jácome (2013), mesmo que o resultado médio seja menor.

Esta afirmação é apoiada pelo argumento de que o design e a aplicação de rubricas aumentam a validade e a transparência na avaliação, promovendo a democracia e um impacto positivo na aprendizagem, culminando em avaliações equitativas, sendo todos esses princípios inerentes à justiça (JÁCOME, 2013).

Isto indica que, em uma análise superficial, o professor tende a sobrepontuar os estudantes segundo a metodologia tradicional, o que não pode ser efetuado com a rubrica. Isto está de acordo com o que postula Moreira (2013), o qual destaca que existe a possibilidade de professores sobrepontuarem os estudantes, inclusive como forma de não permitir que o professor seja avaliado de forma negativa (MOREIRA, 2013).

Ao observarmos as respostas da prova do acadêmico 7, na primeira questão, que solicitava a definição de Promoção da Saúde, utilizando o conceito mais atual, sua resposta foi “Capacitação da comunidade para atuar na melhoria da sua qualidade de vida e saúde”. Analisando-se as notas atribuídas pelos cinco professores que realizaram a correção pelo método tradicional, obtivemos que dois deles atribuíram valor 10 à resposta, 2 deles nota 9, e um deles nota 7. Desse modo, não sabemos as razões que levaram esse professor a optar pela nota 7, não há explicação da correção, o que ele observou de erro e nem as informações que ele julga terem faltado.

A falta do *feedback* pode ser prejudicial para o aprendizado do aluno, tendo em vista que, segundo Duarte et al., (2012), é um elemento fundamental para a melhoria das atividades de ensino e de aprendizagem em que estão envolvidos professores e alunos. O *feedback* pode ser identificado como uma das possíveis estratégias de avaliação para a aprendizagem, dado o poder de fomentar o questionamento, a autoavaliação ou a avaliação por pares.

A mesma questão foi analisada em relação ao acadêmico 32, que respondeu “Através de mecanismos de participação da sociedade e de órgãos públicos em parceria atuando na

promoção de saúde”, na qual, dos cinco professores que corrigiram pelo método tradicional, dois deles atribuíram nota zero à questão, um deles nota 6, nota 5 e o último, nota 4. Nesse caso, podemos observar grande discrepância, principalmente em relação aos professores que atribuíram zero em relação ao que atribuiu 6. Esses fatos vêm ao encontro dos estudos de Erickson (2001), Charney (1984) e Stach (1987), que postulam que as pontuações feitas por vários avaliadores não coincidem nem quanto a erros gramaticais, a vocabulário, a complexidade sintática ou a aparência da resposta.

Outro ponto a ser considerado foi o fato de os professores que realizaram a correção das provas utilizando a rubrica, ou não, terem sido escolhidos de forma aleatória, ou seja, não foi traçado um perfil do profissional levando em consideração seu tempo de atuação como professor, suas áreas de interesse e pesquisa, os métodos de avaliação que utiliza com seus docentes ou sua experiência com a ferramenta *rubric*. Esses fatores poderiam interferir, tendo em vista que os professores com mais tempo de atuação, mais experientes com a disciplina ou habituados com a correção de questões discursivas já podem ter seus próprios critérios de correção bem definidos em mente, o que torna a correção mais padronizada.

Como não foi traçado o perfil do profissional participante previamente e por não haver uma divisão igualitária dos professores que corrigiriam com ou sem a rubrica de acordo com a sua experiência, pode ter ocorrido a concentração de professores mais familiarizados com a disciplina e com a correção de questões dissertativas, o que poderia ser uma explicação para que a diferença entre as correções tenha sido pouco representativa nesse estudo.

8. Considerações finais

O resultado desse trabalho nos permitiu observar que existe uma diferença estatisticamente significativa entre as médias das notas atribuídas pelos professores que realizam a correção de provas dissertativas por meio de rubricas em relação a forma tradicional. No entanto, a diferença de dois pontos em uma escala de zero a 100 foi pouco representativa se analisada em um contexto mais amplo de correção de questões.

No que concerne às limitações do uso das rubricas, podemos destacar que sua elaboração é laboriosa, deve ser construída a partir dos objetivos de aprendizagem e

abranger critérios específicos sem deixar de contemplar itens importantes, sem limitar a avaliação pela falta de critérios. Além disso, um ponto importante seria testá-las antes da aplicação, fornecendo oportunidade para reflexão, opiniões e possibilidade de aperfeiçoamento.

Apesar das limitações, observamos que a utilização de rubricas aumenta a consistência da avaliação, uma vez que indica ao avaliador o que deve ser verificado na correção de questões dissertativas e o nível de proficiência atingido pelo aprendiz. Além disso, um quesito importante na consideração de rubricas em avaliações é a sua visualização prévia pelos estudantes avaliados, de forma que todos possam saber o que se espera com o questionamento abordado, visualizando claramente os critérios considerados no processo de avaliação.

Destaca-se que ainda há um caminho a percorrer para que as rubricas sejam utilizadas com todo o seu potencial, cumprindo seus reais objetivos e sendo aplicadas de forma segura na prática. Porém, vislumbra-se a possibilidade do uso das rubricas como excelente forma de credibilidade e de fidedignidade na correção das questões dissertativas, podendo ser utilizadas em diversas disciplinas, em múltiplos contextos e em variadas áreas de ensino-aprendizagem.

Referências

- BIAGIOTTI, L. C. M. *Conhecendo e aplicando rubricas em avaliações*. Rio de Janeiro: Diretoria de Ensino da Marinha, 2005.
- BLOOM, B.; HASTINGS, J. T.; MADDAUS, G. *Evaluación del aprendizaje*. Buenos Aires: Troquel, 1975.
- BORBA, A. M. de; FERRI, C.; HOSTINS, R. C. L. Avaliação da aprendizagem no ensino superior: questões que emergem da prática docente. *Revista Contrapontos*, Itajaí, SC, v. 7, n. 1, p. 43-54, mar. 2007.
- BROOKHART, S. The art and science of classroom assessment: the missing part of pedagogy. *ASHE-ERIC Higher Education Report*, Trenton, NJ, v. 27, n. 1, p. 112-128, jan. 1999.
- CHARNEY, D. The validity of using holistic scoring to evaluate writing. a critical overview. *Research on the Teaching of English*, Urbana, IL, v. n. 18, p. 65-81, 1984.
- CRUZ, N. K. S.; NUNES, L. C. Delineando rubricas para uma avaliação mediadora da aprendizagem em educação online. [S.l.]: Abed, 2009. Disponível em: <<http://www.abed.org.br/congresso2009/CD/trabalhos/1452009214144.pdf>>. Acesso em: 2 maio 2017.
- DUARTE, P. et al. Avaliação para a aprendizagem em educação a distância: uma revisão integrativa de estudos sobre a utilização de e-rubricas. *TIC e Educação*, [S.l.], 2012. Disponível em: <<http://ticeduca.ie.ul.pt/atas/pdf/6.pdf>>. Acesso em: 4 maio 2017.
- ERICKSON, L. *Stirring the head, heart, and soul: Redefining curriculum and instruction*. Thousand Oaks: Corwin Press, 2001.
- FELICIO, A. C.; CARITÁ, E. C.; OLIVEIRA NETO, J. D. Percepção discente quanto à utilização de rubricas para avaliação da aprendizagem. *Research Gate*, 2013. Disponível em: <https://www.researchgate.net/profile/Jose_Oliveira_Neto/publication/260061194_Percepcao_Discente_Quanto_a_Utilizacao_de_Rubricas_para_Avaliacao_da_Aprendizagem/links/5697dc1f08aea2d74375cc54/Percepcao-Discente-Quanto-a-Utilizacao-de-Rubricas-para-Avaliacao-da-Aprendizagem.pdf>. Acesso em: 2 abr. 2017.
- FERREIRA SOBRINHO, J. W. *Metodologia do ensino jurídico e avaliação em Direito*. Porto Alegre: Sérgio Antonio Fabris Editor, 1997.
- GATICA-LARA, F.; URIBARREN-BERRUETA, T. del N. J. ¿Cómo elaborar una rúbrica?. *Investigacion en Educacion Medica*, Cidade do México, DF, v. 2, n. 1, p. 61-65, jan./mar. 2013.
- HUOT, B. *Validating holistic scoring for writing assessment: theoretical and empirical foundations*. Cresskill: Hampton, 1993.

- JÁCOME, E. P. La rúbrica y la justicia em la evaluación. *Íkala, revista de language y cultura*, Medellín, AT, v. 18, n. 3, p. 79-94, set./dez. 2013.
- LANDIS, J. R.; KOCH, G. The measurement of observer agreement for categorical data. *Biometrics*, Woods Hole, US, v. 33, n. 1, p. 159-174, mar. 1977.
- MENDES, E. A. de M. Avaliação da produção textual nos vestibulares e outros concursos: a questão da subjetividade. *Avaliação*, Sorocaba, SP, v. 18, n. 2, p. 435-458, jul. 2012.
- MONTEIRO, F. de H. Controle jurisdicional da correção de prova subjetiva nos concursos públicos. *Fórum Administrativo*, Belo Horizonte v. 14, n. 164, p. 37-44, out. 2014.
- MOREIRA, D. A. Fatores influentes na avaliação do professor pelo aluno: uma revisão. *Educação e Seleção, Fundação Carlos Chagas*, São Paulo, v. 17, p. 73-87, jan. 2013.
- OLIVEIRA, G. P. Avaliação formativa nos cursos superiores: verificações qualitativas no processo de ensino-aprendizagem e a autonomia dos educandos. *OE – Revista Iberoamericana de Estudos em Educação*, São Paulo, v. 07, n. 1, p. 4-10, mar. 2002.
- PACHECO, J. A avaliação da aprendizagem. In: ALMEIDA, L.; TAVARES, J. (Org.). *Conhecer, aprender e avaliar*. Porto: Porto Editora, 1998, p. 111-132.
- RAPOSO, M.; MARTÍNEZ, E. La Rúbrica en la Enseñanza Universitaria: Un Recurso Para la Tutoría de Grupos de Estudiantes. *Formación universitaria*, Cidade do México, DF, v. 4, n. 4, p. 19-28, maio, 2011.
- STACH, S. C. *The component parts of general impressions: predicting holistic Scores in College level essays*. 1987. 127 f. Dissertação (Mestrado)-University Microfilms International, Arbor, 1987.
- STEIN, L. M. et al. A construção de um instrumento de avaliação discente de um programa de pós-graduação. *Psicologia – USF*, Itatiba, v. 10, n. 02, p. 141-147, jul./ dez. 2005.
- VYGOTSKY, L. S. *A formação social da mente*. São Paulo: Martins Fontes, 1984.
- WOLLCOTT, W. *An overview of writing assessment*. Urbana: National Council of Teachers of English, 1998.
- ZEFERINO, A. M. B.; PASSERI, S. M. R. R. Avaliação da aprendizagem do estudante. *Cadernos Abem*, Goiânia, GO, v. 3, n. 1, p. 39-43, out. 2007.

Recebido em: 16/09/2017

Aceito para publicação em: 12/03/2018

Standardization of the Correction of Dissertative Questions for Collective Health Teachers of the Medicine Course in a Higher Education Institution of the West of Paraná

Abstract

This research refers to the use of essay questions as a method of evaluation and how their correction can vary among different evaluators, resulting in discrepancies in concepts. To avoid such a problem, it is possible to use correction standardization tools, such as rubric. Thus, this study aimed to verify if, with the use of the rubric, the correction becomes clearer and homogeneous. For that, ten teachers of collective health from a college of the West of Paraná were selected to carry out the correction of a dissertation test applied to 33 medical students. Half of these teachers did the correction by the traditional method, while the others corrected using the rubric. After statistical analysis of comparison, it was evidenced that the variability of the means by the traditional correction was greater. In the first questions, the notes are disparate and in the latter, the discrepancy decreases, indicating that, with the traditional method, teachers tend to change the rigor at the end of the correction. In this way, the rubrics can make the correction process more precise and transparent, favoring feedback among teachers and academics.

Keywords: Rubric. Punctuation rubrics. Dissertation proof.

Estandarización de la Corrección de Preguntas Discursivas para Profesores de Salud Colectiva del Curso de Medicina de una Institución de Enseñanza Superior del Oeste del Paraná

Resumen

Este trabajo se refiere al uso de preguntas discursivas como método de evaluación, y cómo su corrección puede variar entre diferentes evaluadores, y traer discrepancias de conceptos. Para evitar tal problemática, se pueden usar herramientas de estandarización de corrección, como la escala descriptiva, denominada rúbrica. Así, el objeto de este estudio fue averiguar si con el uso de las rúbricas, la corrección queda más clara y homogénea. Para ello, se seleccionaron diez profesores de salud colectiva de una IES del

Oeste de Paraná para que realizaran la corrección de una prueba discursiva aplicada a 33 académicos de medicina. La mitad de estos profesores realizó la corrección por el método tradicional, mientras que los demás corrigieron con las rúbricas. Después del análisis estadístico de comparación, quedó claro que la variabilidad de los promedios por la corrección tradicional fue mayor. En los primeros planteos las notas presentan más diferencias y en los últimos la discrepancia disminuye, lo que señala que, con el método tradicional, los profesores tienden a alterar el rigor al final de la corrección. De esta forma, las rúbricas pueden hacer el proceso de corrección más preciso y transparente, favoreciendo la retroalimentación entre profesores y académicos.

Palabras clave: Rúbrica. Rúbricas de calificación. Prueba de carácter disertativo.