

Validação de Instrumento de Avaliação de Disciplinas de Escola Médica no Brasil

Validating a Subject Assessment Tool at a Brazilian Medical School

Margareth Rodrigues Salerno^I
Maria Helena Itaquí Lopes^{II}
Rita Mattiello^I
Edgar Sarria Icaza^{III}
Ivan Carlos Antonello^I

PALAVRAS-CHAVE

- Validação;
- Confiabilidade;
- Validação Convergente;
- Educação Médica;
- Currículo Médico;
- Avaliação;
- Avaliação Docente.

KEYWORDS

- Validation;
- Reliability;
- Convergent Validation;
- Medical Education;
- Medical Curriculum;
- Evaluation;
- Teaching Evaluation

RESUMO

Objetivo: O estudo tem como objetivo avaliar o desempenho do questionário de avaliação das disciplinas da Faculdade de Medicina da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS Famed) em relação às propriedades psicométricas de validade e confiabilidade. *Métodos:* Estudo transversal, exploratório e contemporâneo, com avaliação qualitativa e quantitativa. As disciplinas foram avaliadas pelos alunos de forma anônima, opcional e individual durante um período de aula de grande grupo logo após a conclusão das mesmas. Foram incluídos os questionários respondidos por acadêmicos matriculados entre o primeiro e o oitavo semestre e registrados nas atas finais da unidade, nas 22 disciplinas do currículo, no ano de 2011, que aceitaram realizar a avaliação da disciplina. *Resultados:* Foram estudadas 1.205 respostas de um total possível de 1.606 alunos matriculados. A validação convergente demonstrou forte correlação entre o escore total de satisfação e os campos quantitativos correspondentes. A confiabilidade demonstrou um valor de alfa de Cronbach de 0,753 no escore total. *Conclusão:* O estudo mostrou que o questionário de avaliação de disciplinas é válido e confiável.

ABSTRACT

Aim: The study aims to evaluate a questionnaire on the subjects taught at the Faculty of Medicine of the Pontifical Catholic University of Rio Grande do Sul (PUCRS), in terms of the psychometric properties of validity and reliability. *Methods:* A cross-sectional, exploratory, and contemporary study was conducted using qualitative and quantitative evaluation. The subjects were evaluated by the students in an anonymous, optional and individual fashion during a large group class. The questionnaires were answered by PUCRS academics enrolled on the I-VIII semesters in 2011, who agreed to perform the evaluation immediately after completing the subject and who were registered in the final minutes for the units of the 22 subjects available. *Results:* 1,205 responses were studied out of a possible 1,606 students enrolled on the course. The convergent validation demonstrated a strong correlation between the total score for satisfaction and the quantitative fields. Reliability demonstrated Cronbach alpha from the total score of 0.753. *Conclusion:* The study showed that the questionnaire for subject evaluation is valid and reliable.

Recebido em: 02/07/2015

Aprovado em: 12/01/2016

^I Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brasil.

^{II} Universidade de Caxias do Sul, Caxias do Sul, RS, Brasil.

^{III} Faculdade de Medicina de Santa Cruz do Sul, Santa Cruz, RS, Brasil.

INTRODUÇÃO

Nos últimos anos, as escolas médicas têm buscado desencadear um processo em que professores e alunos participem ativamente do processo de ensino-aprendizagem¹⁻⁴. De forma semelhante, este movimento também tem acontecido no Brasil^{1,5,6}.

No acompanhamento e aprimoramento desse processo, foi necessária a criação de instrumentos que avaliassem o desempenho dessas mudanças propostas nas Diretrizes Nacionais Curriculares para os Cursos de Medicina^{7,8}. O conhecimento dos resultados obtidos permite aprofundar a análise da realidade institucional, rever projetos, adequar metas, corrigir possíveis desvios e diagnosticar fragilidades.

Há questionários nacionais e internacionais validados, utilizados com objetivos amplos e formais de avaliação das instituições de ensino superior^{9,10}. Muitas dessas ferramentas têm como objetivo principal a obtenção de dados globais do ensino nas instituições, não focando a satisfação do aluno em relação aos aspectos gerais e específicos das disciplinas oferecidas durante a sua formação.

A satisfação dos alunos, quando revisada na literatura, relaciona-se ao curso integralmente, explorando desde aspectos de infraestrutura, corpo docente e regime de trabalho, estratégias de ensino, didática em sala de aula, até avaliação do professor, diferentes formas de avaliação em provas e testes, abordando aspectos clínicos e de habilidades⁹⁻¹².

Na Faculdade de Medicina da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), desde 1998, foi desenvolvido um instrumento de avaliação do nível de satisfação com as disciplinas e com o processo ensino-aprendizagem. Trata-se de um questionário que propõe aferir a satisfação do aluno em relação aos diferentes domínios relacionados ao ensino, com impressão geral da disciplina, de vários aspectos específicos da mesma, de forma quantitativa e com espaço aberto, livre para expressão qualitativa. Dessa forma, ao avaliar suas disciplinas de maneira ativa, o aluno participa da elaboração da versão seguinte do currículo da sua escola e, em especial, da discussão dos processos pedagógicos. Adquire condições de olhar de forma mais crítica e criteriosa para sua ação discente, tornando-se responsável e coautor do processo ensino-aprendizagem¹³⁻¹⁵. Aos docentes e à instituição, esse instrumento permite diagnosticar a situação disciplinar do curso e detectar fragilidades, potencialidades e pontos fortes.

Atualmente, no Brasil, não existem instrumentos validados que avaliem, de maneira específica, o nível de satisfação dos alunos em relação ao curso de Medicina. Embora o questionário de avaliação em uso na Famed PUCRS já tenha se mostrado de muita utilidade quanto à veracidade de suas informações, ajudando, muitas vezes, em ajustes no andamento

das disciplinas, é necessário que o instrumento seja avaliado quanto à sua validação. Essa avaliação consiste basicamente no estudo de duas propriedades psicométricas centrais, que podem ser subdivididas em vários elementos nem sempre somatórios¹⁶⁻¹⁹.

Sendo assim, o presente estudo analisou o desempenho do questionário de avaliação das disciplinas da Famed PUCRS em relação às propriedades psicométricas validade e confiabilidade.

MATERIAL E MÉTODO

O presente estudo tem desenho transversal, exploratório e contemporâneo, com avaliação qualitativa e quantitativa.

Todas as variáveis incluídas no presente estudo foram aferidas no mesmo momento do curso. Foram incluídos questionários dos acadêmicos da Famed PUCRS que estavam cursando entre o primeiro e o oitavo semestre no ano de 2011, que aceitaram realizar a avaliação logo após a conclusão da disciplina e estavam registrados nas atas de aproveitamento final das 22 disciplinas disponibilizadas. Foram excluídos os questionários daqueles alunos que entregaram o instrumento em branco ou com rasuras que comprometessem o entendimento do que foi preenchido, ou o conjunto de questionários de disciplinas que apresentaram um número total de respondentes menor que 20% dos acadêmicos matriculados.

Questionário de avaliação do nível de satisfação dos alunos

O instrumento teve origem no interesse da Direção da escola médica da PUCRS, em 1998, de conhecer o nível de satisfação de seus alunos com o curso de graduação e com o potencial de informações que poderia ser obtido numa avaliação do curso a partir dos alunos. Um grupo de professores da Faculdade de Medicina e da Faculdade de Educação da PUCRS desenvolveu esse instrumento interno específico para avaliação das disciplinas. A principal vantagem da construção desse questionário era que seu objetivo principal consistia em responder às peculiaridades e necessidades das escolas de Medicina. O grupo foi formado por professores e representantes do Diretório Acadêmico da Famed.

Após várias revisões de literatura e discussões em grupo, a primeira versão do instrumento foi testada entre os professores da escola e depois em um grupo de alunos. O resultante deste estudo piloto é o instrumento atual, que vem sendo utilizado regularmente nas disciplinas da Famed desde esse período. Esta ferramenta, apesar da experiência dos docentes e da qualidade das informações obtidas, carecia de análise e validação a partir dos resultados, o que originou este estudo.

Perguntas objetivas

O questionário de avaliação das disciplinas da Famed PUCRS é constituído de dez questões. As opções de respostas do item impressão geral são em escala Likert, composta por cinco opções que variam desde ótimo até péssimo. As opções de respostas dos demais itens são: “sim”, “não” e “não se aplica”. A avaliação finaliza com um espaço aberto para comentários livres. A seguir, apresentamos as questões que compõem o instrumento:

- Qual a impressão geral sobre a disciplina?
- O plano de ensino foi apresentado aos alunos?
- A orientação bibliográfica como fonte de consulta foi satisfatória (atualizada e disponível)?
- A disciplina possibilitou atividades práticas?
- Os objetivos propostos pelo plano de ensino foram atingidos?
- As provas foram bem elaboradas?
- A avaliação do aluno foi adequada ao plano de ensino apresentado?
- Os resultados da avaliação foram discutidos com os alunos?
- A carga horária total foi cumprida e bem aproveitada?
- Houve integração com outras disciplinas do currículo?

As questões quantitativas foram agrupadas em domínios, conforme sua afinidade, gerando quatro domínios: impressão geral, dinâmica da disciplina, avaliação e matriz curricular.

Comentários livres

Os comentários livres foram estudados usando-se a metodologia de análise de conteúdo²⁰, conforme proposição de Bardin, segundo a qual qualquer forma de comunicação, seja escrita ou dita, é passível de análise de conteúdo²¹. Esse método, que analisa material subjetivo, atende ao rigor científico, pois não se propõe a uma leitura simples da realidade. A análise de conteúdo se desenvolve em etapas sucessivas, compreendendo três fases: pré-análise, exploração do material e tratamento dos resultados.

Sistemática utilizada para interpretar e classificar as opiniões dos alunos em categorias

Categoria professores: quando o aluno fazia referência nominal a algum professor, ao modo como o professor ministrou a aula, a alguma atitude ou comportamento durante a aula (controle sobre os alunos, listas de presenças, postura em sala de aula), referência a substitutos.

Categoria dinâmica da disciplina: sempre que o aluno referia algo em relação ao formato da aula, organização, ho-

rários, permanência dos professores, trocas de programação, número de alunos por turma, repetição de assuntos e satisfação com a disciplina em si.

Categoria avaliação da disciplina: toda citação que se referia a notas, elaboração de provas, momento de realização de provas, correção de provas, comparações de notas, referenciais de avaliação, discussão de provas ou de avaliação.

Categoria matriz curricular: quando o aluno se referia à posição da disciplina no currículo ou qualquer sugestão a esse respeito, integração das disciplinas, referência ao currículo anterior.

Categoria infraestrutura: sempre que o aluno se referia às condições físicas para o desenvolvimento das atividades, aos materiais utilizados, aos ambientes comuns da faculdade.

As disciplinas foram avaliadas pelos alunos de forma anônima, opcional e individual, durante um período de aula no final do semestre.

As categorias qualitativas foram agrupadas em domínios, conforme a afinidade.

Para analisar os resultados e aferir a validade do instrumento, foram, então, utilizados os domínios quantitativos e qualitativos formados.

Formação de domínios quantitativos

Os itens quantitativos foram agrupados em domínios, conforme sua afinidade, gerando quatro domínios: impressão geral, dinâmica da disciplina, avaliação e matriz curricular.

Desta forma, foram estabelecidos cinco domínios da parte qualitativa: professores, dinâmica da disciplina, avaliação, matriz curricular e infraestrutura, separados em positivos, negativos, e positivos mais negativos.

Validação do questionário

Para realizar a validação do instrumento, as propriedades psicométricas estudadas foram validade e confiabilidade. A validade foi avaliada por meio da validade convergente, e a confiabilidade, mediante a consistência interna e os efeitos teto e chão^{16,18,19}.

Na validade convergente, o instrumento deve se correlacionar com as informações com as quais o construto deveria estar relacionado, demonstrando que o instrumento é válido para avaliar o que se pretende. A validade convergente foi avaliada por meio das correlações do escore total com os domínios quantitativos e desses entre si. Adicionalmente, avaliou-se a diferença entre as médias dos domínios quantitativos e qualitativos correspondentes (para o domínio dinâmica da disciplina, que foi o domínio com informações suficientes para essa análise).

A confiabilidade, que avalia se um instrumento é capaz de medir sempre da mesma maneira o que se pretende medir, foi estudada por meio da consistência interna, utilizando-se o coeficiente alfa de Cronbach ($\alpha - C$). Os efeitos teto e chão foram calculados a partir da percentagem de respostas, respectivamente, sim (efeito teto) e não (efeito chão) do instrumento.

O efeito teto é a percentagem da população que pontua no mais elevado nível da medida. O efeito chão, por sua vez, se manifesta quando a percentagem dos sujeitos pontua no mais baixo nível da medida^{18,19,22}.

Cálculo amostral

Para a avaliação das propriedades psicométricas confiabilidade e validade, o cálculo amostral foi de 300 participantes, tendo em vista a inclusão de 30 sujeitos para cada item do instrumento^{16,19}.

Quanto ao tratamento estatístico, as variáveis categóricas foram descritas por frequências absolutas e relativas, e as variáveis contínuas foram descritas mediante médias e desvio padrão ou mediana e intervalo interquartil, conforme a distribuição das variáveis.

A validade convergente do instrumento foi analisada de duas formas: (a) associação entre escore total e os domínios quantitativos e desses entre si; (b) comparação entre as médias dos escores dos domínios quantitativos com as médias dos escores dos domínios qualitativos (domínio dinâmica da disciplina). A comparação entre as médias dos escores dos domínios quantitativos com as respostas do grupo de domínios qualitativos foi avaliada mediante o teste de Anova. Para avaliar a diferença nas médias do domínio dinâmica da disciplina, utilizamos o *post hoc* teste de Tukey. Já as possíveis relações entre escore total e os domínios quantitativos foram estudadas por meio da correlação de Spearman (r), e consideramos aceitáveis as correlações com valores de $r > 0,2$ ²².

A confiabilidade foi testada mediante a consistência interna e percentagem dos efeitos teto e chão. Para avaliarmos a consistência interna do instrumento, utilizamos o Coeficiente Alfa de Cronbach ($\alpha - C$), considerando adequados os valores $\geq 0,6$ ^{16,18}.

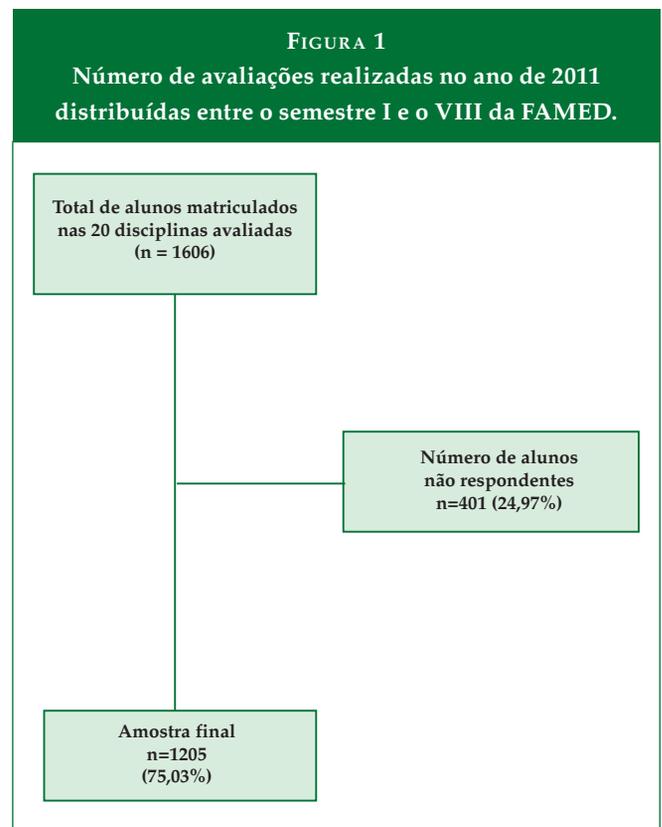
Todos os testes foram bidirecionais, e as diferenças consideradas significativas com $p < 0,05$. Utilizou-se o Excel for Windows 2010 para organização das informações em um banco de dados, e as análises foram realizadas com o programa SPSS versão 17.0 (SPSS Inc., Chicago IL).

A pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da PUCRS em 7 de outubro de 2011, com o protocolo de pesquisa registro CEP 11/05600.

RESULTADOS

Do total de 22 disciplinas do currículo, de responsabilidade administrativa da Famed, 20 foram avaliadas no estudo. As duas disciplinas não avaliadas pelos alunos foram Aproximação à Prática Médica III (semestre 3) e Epidemiologia Geral (semestre 5), perdidas do seguimento por apresentarem menos de 20% dos questionários possíveis respondidos.

O total de alunos matriculados nas 20 disciplinas foi de 1.606. Desses, 1.205 (75%) responderam, de forma anônima e opcional, aos questionários de avaliação durante o ano de 2011 (Figura 1).



As disciplinas estavam distribuídas do primeiro ao oitavo semestre do curso. Dos 1.205 questionários respondidos, a maioria, 688 (57%), fez também comentários livres para sua comunicação.

Ao avaliarmos os domínios quantitativos do questionário, encontramos que as médias dos domínios oscilaram de 77,34 (domínio impressão geral) a 87,92 pontos (dinâmica da disciplina). Quanto à distribuição do escore total, observamos que a média do escore total foi igual ou superior a 85 pontos na maioria dos casos, 851 (70,6%) (Tabela 1).

TABELA 1
Médias dos domínios quantitativos

Domínio	Média	Desvio-Padrão
Impressão geral	77,34	± 17,34
Dinâmica da disciplina	87,92	± 17,75
Avaliação	80,86	± 27,43
Matriz curricular	84,30	± 36,40
Escore total	84,38	± 36,40

Os comentários livres estavam presentes em 688 (57%) questionários, tendo sido categorizados mediante análise de conteúdo. Os resultados são demonstrados na Tabela 2.

TABELA 2
Distribuição dos comentários livres por domínios

Domínios Qualitativos	n	(%)
	1.205	
Professores		
Sem comentários	979	81,2
Comentários (positivos)	50	4,1
Comentários (negativos)	147	12,2
Comentários (negativos e positivos)	29	2,4
Dinâmica da disciplina		
Sem comentários	690	57,3
Comentários (positivos)	111	9,2
Comentários (negativos)	304	25,2
Comentários (negativos e positivos)	100	8,3
Avaliação		
Sem comentários	988	82
Comentários (positivos)	7	0,6
Comentários (negativos)	206	17,1
Comentários (negativos e positivos)	4	0,3
Matriz curricular		
Sem comentários	1.111	92,2
Comentários (positivos)	2	0,2
Comentários (negativos)	92	7,5
Comentários (negativos e positivos)	0	0
Infraestrutura		
Sem comentários	1.143	94,9
Comentários (positivos)	3	0,2
Comentários (negativos)	59	4,9
Comentários (negativos e positivos)	0	0

n = número absoluto.

Em relação aos comentários livres, houve predomínio de comentários negativos em todos os domínios. Os domínios avaliação, 206 (17%), e dinâmica da disciplina, 304 (25%), foram os mais referidos de forma negativa pelo maior percentual de alunos.

Quanto à dinâmica da disciplina, em 15 (75%) delas, encontrou-se que mais de 20% dos alunos por disciplina fizeram comentários negativos. Em 19 (95%) das disciplinas estudadas, os alunos se referiram ao domínio avaliação de forma negativa, sendo que em apenas uma (Saúde da Criança e do Adolescente) não fizeram qualquer comentário sobre avaliação. Na Prática em Saúde do Adulto e do Idoso I, a avaliação não foi considerada, pois o questionário foi respondido previamente. Mesmo em relação à disciplina como segundo melhor escore total (Trauma e Emergência), 50,7% fizeram comentários negativos sobre as práticas avaliativas da mesma. Do universo total de disciplinas, 65% receberam críticas negativas em relação ao domínio matriz curricular. Com referência à infraestrutura oferecida, menos de 10% fizeram qualquer observação.

Validade convergente

Ao correlacionamos os domínios quantitativos entre si e com o escore total, observamos que a maioria das associações avaliadas foi aceitável e significativa. O escore total apresentou uma associação expressiva e significativa com todos os domínios avaliados (Tabela 3).

TABELA 3
Associação entre os domínios quantitativos e escore total

	Dinâmica da disciplina	Avaliação	Matriz curricular	Escore total
	rP	r P	r P	r P
Impressão geral	0,475* < 0,001*	0,362* < 0,001*	0,321* < 0,001*	0,709* < 0,001*
Dinâmica da disciplina		0,346* < 0,001*	0,278* < 0,001*	0,767* < 0,001*
Avaliação			0,258* < 0,001*	0,742* < 0,001*
Matriz curricular				0,476* < 0,001*

* A associação entre os domínios foi avaliada pelo Teste de Spearman. As correlações foram consideradas significativas a partir dos valores de $P < 0,05$.

Esses resultados contribuem para a validação do construto como planejado, mostrando que o escore total apresentou boa correlação com os domínios quantitativos, impressão geral, dinâmica da disciplina e avaliação, e moderada correlação com a matriz curricular (Tabela 3). Adicionalmente, no domínio dinâmica da disciplina, o único com número de respondentes suficientes para análise, houve correlação entre as categorias qualitativas e a pontuação quantitativa de satisfação (Tabela 4).

As médias do domínio dinâmica da disciplina (dados quantitativos) apresentaram diferenças significativas entre as categorias da parte qualitativa do questionário. As diferen-

ças entre as médias dos domínios quantitativos “avaliação” e “matriz curricular” e as categorias dos domínios qualitativos correspondentes não foram avaliadas devido ao número limitado de comentários livres nesses domínios.

Dinâmica da disciplina (qualitativo) Categorias	Dinâmica da disciplina (quantitativo) Média ± DP	P
Sem comentários (n = 690)	90,06 ± 16,01	< 0,001*
Comentários (positivos) (n = 111)	96,73 ± 8,36	
Comentários (negativos) (n = 304)	82,13 ± 20,80	
Comentários (negativos e positivos) (n = 100)	85,81 ± 19,24	

* As médias das categorias foram comparadas mediante Análise de Variância (Anova). Foram consideradas como diferenças significativas os valores $P < 0,001$.

Como mostra a Tabela 4, os alunos que fizeram comentários positivos tiveram maior pontuação no escore do questionário quantitativo, quando comparados com os que não fizeram comentários ($p = 0,001$), com os que fizeram comentários negativos ($p < 0,001$) e com os que fizeram comentários positivos e negativos ($p < 0,001$).

Os alunos que não fizeram comentários também apresentaram maior pontuação quando comparados com os que fizeram comentários negativos ($p < 0,001$) (Tabela 4).

Confiabilidade

O alfa de Cronbach ($\alpha - C$) do escore total foi de 0,753.

Todos os domínios apresentaram o efeito chão inferior a 20%. O efeito teto apresentou valor superior a 20% em todos os domínios avaliados.

Os resultados das propriedades psicométricas sugerem que o questionário de avaliação da satisfação dos alunos da Famed é válido e confiável.

RESULTADOS GERAIS

A Tabela 5 apresenta o número de alunos matriculados, o número de questionários respondidos e o escore total das disciplinas nos domínios quantitativos.

A média do escore total das perguntas quantitativas é de 84,38 pontos (Tabela 1). Quando avaliada por disciplina (Tabela 5), encontrou-se que 9 (45%) disciplinas têm escore total acima de 85 pontos.

TABELA 5
Disciplinas avaliadas, número de estudantes matriculados, número de questionários respondidos e escore total dos domínios quantitativos

Disciplinas avaliadas	Número de alunos matriculados, n (%)	Número de questionários respondidos, n (%)	Escore total, média ± desvio-padrão
Aproximação à Prática Médica I	75	56 (75)	97,5 ± 4,08
Aproximação à Prática Médica II	74	67 (91)	90,44 ± 9,53
Mecanismo de Doença I	80	57 (71)	95,24 ± 7,81
Mecanismo de Doença II	79	69 (87)	95,52 ± 7,35
Mente e Cérebro I	81	56 (69)	75,39 ± 18,13
Saúde Coletiva I	75	72 (96)	80,18 ± 17,93
Propedêutica I	79	60 (76)	76,08 ± 17,78
Mente e Cérebro II	88	66 (75)	89,11 ± 11,64
Saúde Coletiva II	89	87 (98)	69,54 ± 17,81
Propedêutica II	90	61 (68)	76,59 ± 14,24
Prática em Saúde do Adulto e do Idoso I	88	25 (28)	*
Saúde da Criança e do Adolescente	87	87 (100)	95,04 ± 7,14
Laboratório de Habilidades I	88	62 (70)	89,73 ± 9,42
Prática em Saúde do Adulto e do Idoso II	78	41 (53)	84,72 ± 12,06
Saúde Materno-Infantil	77	48 (62)	78,80 ± 17,95
Prática em Saúde do Adulto e do Idoso III	76	62 (82)	90,34 ± 10,21
Laboratório de Habilidades II	77	26 (34)	80,73 ± 13,15
Saúde da Mulher e Materno-Infantil	76	62 (82)	82,27 ± 15,34
Trauma e Emergências	74	71 (96)	96,25 ± 7,23
Medicina Legal e Deontologia Médica	75	70 (93)	65,17 ± 19,24

* O domínio disciplina não foi avaliado nesse momento, pois o questionário foi aplicado antes da avaliação final.

DISCUSSÃO

O questionário de avaliação das disciplinas da Famed PUCRS apresentou bom desempenho em relação às propriedades psicométricas validade e confiabilidade.

No estudo, a confiabilidade foi adequada tanto na correlação entre os itens do questionário quantitativo entre si e com o escore total, quanto no coeficiente alfa de Cronbach, demonstrando que o questionário apresenta consistência interna adequada. Dessa forma, obtivemos que o instrumento está adequado quanto à propriedade central confiabilidade^{16-19,23,24}.

A avaliação da validade convergente foi realizada por meio das correlações do escore total e dos domínios quantitativos e mediante a diferença entre as médias dos domínios

quantitativos com os qualitativos correspondentes. As correlações entre os domínios e o escore total apresentaram valores aceitáveis e significativos. Os resultados demonstraram que as respostas quantitativas e qualitativas convergiram na mesma direção, confirmando a propriedade de validade convergente do instrumento.

As associações adequadas e o valor do alfa de Cronbach refletem que a elaboração do instrumento incluiu perguntas que convergem para o mesmo propósito e que avaliam o que se propõem avaliar.

O estudo demonstrou que a média de respondentes ao questionário (78%) foi similar à média descrita na literatura, que refere de 49% a 84%^{25,26,27}. O número adequado de respondentes pode ser considerado um indicador de confiança nos resultados²⁸.

Existe um número limitado de instrumentos validados de avaliação sobre o nível de satisfação dos alunos em relação ao curso. Na área da saúde, diferentemente de áreas como Psicologia, ciências educacionais, sociais e jurídicas, que lidam com métodos psicométricos de validação desde a década de 1940¹⁶, essa prática ainda apresenta dificuldade pela falta de familiaridade com a operacionalização de variáveis subjetivas. Provavelmente, esta seria uma das possíveis explicações para a falta de questionários validados no País.

O questionário de avaliação das disciplinas em uso na Famed PUCRS, até agora, vem se mostrando de muita utilidade quanto à veracidade de suas informações, ajudando em ajustes no andamento das disciplinas, e tem sido fundamental nas reorganizações curriculares. No entanto, era necessário avaliar o instrumento quanto à sua validação, uma vez que esse processo atribui qualidade ao instrumento de medição.

Em relação ao nível de satisfação dos alunos com as disciplinas da Famed no ano de 2011, o estudo mostrou que a maior parte (70,6%) se manifestou satisfeita, conforme o escore total de satisfação. Quando se observa a avaliação global, o estudo mostra que existem reparos pontuais que necessitam ser feitos. No entanto, quando comparamos esses valores com os da literatura, o percentual de satisfação dos alunos nos diferentes cursos é semelhante ao encontrado no estudo. Marcondes e Lima Gonçalves²⁸, na Universidade de São Paulo, mostram satisfação de 70% a 80% em diferentes blocos de disciplinas. Ruiz e Morita²⁵, em Botucatu, encontraram 47,7% de conceito excelente em relação ao currículo da escola. Clack¹³, no Kings College School of Medicine and Dentistry of London, encontrou 70% de satisfação entre os alunos. Na Hebrew University de Jerusalém, em Israel, Eyal *et al.*²⁶ encontraram que, apesar de apenas 60% dos respondentes entre alunos e médicos já graduados estarem satisfeitos com sua educação

médica, 33% concordavam em que os conhecimentos fundamentais, habilidades, atitudes e valores que todo médico deve ter na sua graduação tinham ficado explicitamente claros para todos quando estudantes. Na Linköping University, Faculty of Health Sciences, na Suécia, Antepohl *et al.*²⁹ estudaram a satisfação dos alunos da graduação após a implantação de um novo currículo, tendo obtido alto grau de satisfação dos egressos quanto à sua educação na graduação e à sua preparação para a prática médica.

Utilizamos apenas o domínio dinâmica da disciplina por ter sido o único com número de respostas que permitem poder estatístico para a análise proposta. No entanto, a análise dos comentários livres se correlacionou com a análise dos dados quantitativos, o que agregou melhor entendimento da opinião dos alunos, com enfoques complementares, e conferiu mais confiança aos resultados. Wassermann *et al.*¹⁵ também encontraram concordância nas opiniões dos alunos sobre seu ensino, do ponto de vista qualitativo e quantitativo. Avaliamos as respostas dos comentários livres com base na reflexão de Cashin¹¹. Consideramos relevantes todos os comentários em que mais de dez alunos por disciplina se manifestaram.

Uma das limitações do estudo é a de não se ter caracterizado a população avaliada, em função da necessidade de anonimato. No entanto, vários autores, entre eles Cashin¹¹ e McKeachie³⁰, afirmam que a idade, o sexo e a posição no curso não modificam os resultados nas avaliações. Já o anonimato é necessário para um resultado adequado na medida em que aumenta a segurança do respondente e a veracidade das respostas³¹⁻³³.

Pode-se considerar também uma restrição o fato de que as propriedades psicométricas reprodutibilidade e sensibilidade à mudança não foram avaliadas. Como o questionário foi construído para avaliar o nível de satisfação dos alunos em relação ao ano letivo em curso, a reprodutibilidade e a sensibilidade apenas poderiam ter sido avaliadas se houvesse uma proposta de mudanças no plano de ensino e se o questionário tivesse sido aplicado antes e após essa intervenção.

O fato de termos trabalhado com opções de respostas categóricas também é uma limitação, uma vez que não permitiu dados intermediários, apenas antagônicos. Isso nos traz a perspectiva de melhorar o instrumento, transformando os nove itens quantitativos do instrumento interno de respostas bimodais em respostas em escalas Likert.

O questionário revelou que o efeito teto era superior a 20% em todos os domínios avaliados. Embora os dados tenham demonstrado que a maioria dos alunos estava satisfeita com as disciplinas, tivemos um número expressivo de comentários livres negativos, o que sugere que os alunos participaram de

maneira ativa da avaliação. Com maior número de opções, numa escala Likert, talvez seja possível melhor distribuição de respostas.

CONCLUSÃO

O estudo demonstrou que este instrumento interno de avaliação da Famed PUCRS é confiável e válido, de acordo com as propriedades psicométricas estudadas.

REFERÊNCIAS

- Sandars J, Hart Sarojin C. The capability approach for medical education: AMEE Guide No. 97. *Medical Teacher*. 2015; 37(6): 510-520.
- Stigler FL, Duvivier RJ, Weggemans M, Salzer HJ. Health professionals for the 21st century: a students' view. *Lancet*. 2010; 376(9756):1877-8.
- Cooke M, Irby DM, Sullivan W, Ludmerer KM. American medical education 100 years after the Flexner report. *N Engl J Med*. 2006;355(13):1339-44.
- Gomes AP, Rego S. Transformação da educação médica: é possível formar um novo médico a partir de mudanças no método de ensino-aprendizagem. *Rev Bras Edu Méd*. 2011;35(4):557-66.
- Rêgo C, Batista SH. Faculty development courses in medicine: a fruitful field. *Revista Brasileira de Educação Médica*. 2012;36(3):317-24.
- Lampert JB, Aguilar-da-Silva R, Perim G, Stella R, Abdalla I, Campos Costa N. Projeto de avaliação de tendências de mudanças no curso de graduação nas escolas médicas brasileiras. *Rev Bras Educ Med*. 2009;33(Supl 1):5-18.
- Brasil. Ministério da Educação. MEC. CES nº4 de 7 de novembro de 2001: Institui diretrizes curriculares nacionais do curso de graduação em Medicina Resolução CNE. CES nº4 de. 2001;7.
- Guimarães RGM, Ferreira MC, Villaça FM. O Debate Necessário: a importância da extensão universitária para a formação médica. *Cadernos ABEM*. 2008;4:69-78.
- Bittencourt HR, Creutzberg M, de Moraes Rodrigues AC, de Oliveira Casartelli A, de Freitas ALS. Desenvolvimento e validação de um instrumento para avaliação de disciplinas na educação superior. *Estudos em Avaliação Educacional*. 2013;22 (48):91-113.
- Keane E, Labhrainn I. Obtaining student feedback on teaching & course quality. *Briefing paper*. 2005;2:1-19.
- Cashin WE. Student Ratings of Teaching: The Research Revisited. *IDEA Paper No. 32*. 1995.
- INEP. Disponível em: < <http://www.inep.gov.br>>. Acesso em Mai 2013. 2011.
- Clack GB. Medical graduates evaluate the effectiveness of their education. *Med Educ*. 1994;28(5):418-31.
- Gerbase MW, Germond M, Nendaz MR, Vu NV. When the evaluated becomes evaluator: what can we learn from students' experiences during clerkships? *Acad Med*. 2009;84(7):877-85.
- Wassermann A, Harris M, Demeco S, Bernabó J. Un sistema para hallar oportunidades de mejora en la enseñanza basado en la percepción de los alumnos. *Educación médica*. 2006;9(2):66-74.
- Pasquali L. *Psicometria: teoria dos testes na psicologia e na educação*: Vozes; 2009.
- Speyer R, Pilz W, Van Der Kruis J, Brunings JW. Reliability and validity of student peer assessment in medical education: a systematic review. *Med Teach*. 2011;33(11):e572-85.
- Pasquali L. *Psychometrics*. *Revista da Escola de Enfermagem da USP*. 2009;43(SPE):992-9.
- Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. *Med Care*. 2000;38(9 Suppl):II84-90.
- Moraes R. Análise de conteúdo. *Revista Educação, Porto Alegre*. 1999;22(37):7-32.
- Bardin L. Análise de conteúdo. Edição revista e atualizada. Lisboa: Edição 2009. 48 p.
- Ebel RL. *Essentials of educational measurement*. 1972.
- Haynes SN, Richard D, Kubany ES. Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological assessment*. 1995;7(3):238.
- Icaza EES. Validação de campo dos questionários de qualidade de vida relacionada à saúde, o paediatric asthma quality of life questionnaire e o paediatric quality of life inventory em crianças asmáticas do Rio Grande do Sul. 2007.
- Ruiz T, Morita I. Curso de graduação na Faculdade de Medicina de Botucatu-UNESP: inquérito entre ex-alunos. *AMB rev Assoc Med Bras*. 1991;37(4):200-4.
- Eyal L, Cohen R. Preparation for clinical practice: a survey of medical students' and graduates' perceptions of the effectiveness of their medical school curriculum. *Med Teach*. 2006;28(6):e162-70.
- Metcalfe DH, Matharu M. Students' perception of good and bad teaching: report of a critical incident study. *Med Educ*. 1995;29(3):193-7.
- Marcondes E, Lima Gonçalves E. Programa de avaliação curricular da Faculdade de Medicina da Universidade de São Paulo. *Rev Hosp Clin Fac Med Univ São Paulo*. 1991;46(5):243-9.
- Antepohl W, Domeij E, Forsberg P, Ludvigsson J. A follow-up of medical graduates of a problem-based learning curriculum. *Med Educ*. 2003;37(2):155-62.

30. McKeachie WJ. Student ratings: The validity of use. 1997.
31. Woloschuk W, Coderre S, Wright B, McLaughlin K. What factors affect students' overall ratings of a course? *Acad Med.* 2011;86(5):640-3.
32. Schönrock-Adema J, Lubarsky S, Chalk C, Steinert Y, Cohen-Schotanus J. 'What would my classmates say?' An international study of the prediction-based method of course evaluation. *Med Educ.* 2013;47(5):453-62.
33. Raupach T, Münscher C, Beissbarth T, Burckhardt G, Pukrop T. Towards outcome-based programme evaluation: using student comparative self-assessments to determine teaching effectiveness. *Med Teach.* 2011; 33(8):e446-53.

CONTRIBUIÇÃO DOS AUTORES

Margareth Rodrigues Salerno: Elaboração do projeto, coleta dos dados, análise dos dados, escrita do artigo. **Maria Helena Itaquí Lopes:** Elaboração do projeto, análise dos dados, escrita do artigo. **Rita Mattiello:** Elaboração do projeto, análise dos dados, escrita do artigo. **Edgar Sarria Icaza:** Elaboração do projeto, análise dos dados, escrita do artigo. **Ivan Carlos Antonello:** Elaboração do projeto, análise dos dados, escrita do artigo.

CONFLITO DE INTERESSES

Os autores declaram não haver conflito de interesses.

ENDEREÇO PARA CORRESPONDÊNCIA

Margareth Salerno
Hospital São Lucas da PUCRS
Av. Ipiranga, 6690 – Centro Clínico – sala 420
CEP 90 610-000 – RS
E-mail: margareth.salerno@pucrs.br