

Desenvolvimento e validação de um instrumento para avaliação de disciplinas na educação superior

HÉLIO RADKE BITTENCOURT*

MARION CREUTZBERG**

ALZIRO CÉSAR DE MORAIS RODRIGUES***

ALAM DE OLIVEIRA CASARTELLI****

ANA LÚCIA SOUZA DE FREITAS*****

RESUMO

A avaliação educacional e, mais especificamente, a avaliação de disciplinas geram polêmicas no âmbito da educação superior, a começar pelos instrumentos utilizados. Este artigo tem por objetivo apresentar as etapas do processo de desenvolvimento e de validação de um instrumento de avaliação para disciplinas de graduação, utilizando técnicas oriundas da Psicometria. Elaborou-se um instrumento simples e pouco extenso que contemplasse três dimensões específicas: 1) aspectos gerais da disciplina; 2) ação de ensino do professor; e 3) autoavaliação

* Professor da Faculdade de Matemática da Pontifícia Universidade Católica do Rio Grande do Sul (heliorb@pucrs.br).

** Professora da Faculdade de Enfermagem, Nutrição e Fisioterapia da Pontifícia Universidade Católica do Rio Grande do Sul (mcreutzberg@pucrs.br).

*** Professor da Faculdade de Administração, Contabilidade e Economia da Pontifícia Universidade Católica do Rio Grande do Sul (rodrigues@pucrs.br).

**** Professor da Faculdade de Administração da Pontifícia Universidade Católica do Rio Grande do Sul (alam@pucrs.br).

***** Professora da Faculdade de Educação da Pontifícia Universidade Católica do Rio Grande do Sul (ana.freitas@pucrs.br).

(ação do aluno). Dentre as técnicas utilizadas no processo de validação, destacam-se a realização de reuniões com especialistas e o uso de técnicas estatísticas multivariadas de Análise Fatorial Confirmatória e Análise de Regressão. Os resultados obtidos com base em um conjunto de 151.220 instrumentos, preenchidos por 32.124 estudantes em três períodos letivos, revelaram que o instrumento proposto apresenta qualidades psicométricas desejáveis relativamente à sua validação e fidedignidade. Adicionalmente, sustenta-se que a avaliação de disciplinas não seja vista como um processo desconectado do projeto pedagógico institucional, contando com a participação de profissionais de diferentes áreas no desenvolvimento do instrumento e na condução do processo, com base em uma perspectiva de avaliação formativa, orientada à promoção da aprendizagem dos estudantes.

Palavras-chave: Ensino superior, Avaliação da educação, Psicometria, Avaliação da aprendizagem.

RESUMEN

La evaluación educativa y, más específicamente, la evaluación de asignaturas generan controversias en el ámbito de la educación superior, comenzando por los instrumentos utilizados. Este artículo tiene como objetivo presentar las etapas del proceso de desarrollo y de validación de un instrumento de evaluación para asignaturas de grado, utilizando técnicas de la Psicometría. Se elaboró un instrumento sencillo y poco extenso que contemplase tres dimensiones específicas: 1) aspectos generales de la asignatura; 2) la acción del docente y 3) la autoevaluación (acción del alumno). Entre las técnicas utilizadas en el proceso de validación se destaca la realización de reuniones con especialistas y el uso de técnicas estadísticas multivariadas de Análisis Factorial Confirmatorio y Análisis de Regresión. Los resultados obtenidos a partir de un conjunto de 151.220 instrumentos, llenados por 32.124 estudiantes en tres períodos lectivos, revelaron que el instrumento propuesto presenta cualidades psicométricas deseables en relación a su validación y confiabilidad. Además, se argumenta que la evaluación de asignaturas no sea vista como un proceso desconectado del proyecto pedagógico institucional. El mismo cuenta con la participación de profesionales de diferentes áreas en el desarrollo del instrumento y en la conducción del proceso, con base en una perspectiva de evaluación formadora, orientada a la promoción del aprendizaje de los estudiantes.

Palabras clave: Educación superior, Evaluación de la educación, Psicometría, Evaluación del aprendizaje.

ABSTRACT

The issue of educational evaluation and, more specifically, the evaluation of courses generates controversy in the context of higher education, chiefly considering the instruments adopted. This article presents the stages of development and validation of an evaluation instrument for undergraduate courses, using Psychometric techniques. A simple and not extensive instrument was created to cover three specific dimensions: 1) general aspects of the course, 2) teacher's teaching action, and 3) self-evaluation (student's action). Among the techniques used in the validation process we highlight interviews with experts and the use of multivariate techniques

of Confirmatory Factor Analysis and Regression Analysis. The results obtained from a set of 151,220 instruments, completed by 32,124 students in three academic periods, revealed that the instrument has desirable psychometric qualities as regards its reliability and validity. Additionally, it is argued that evaluation of courses should not be seen as a process disconnected from the institutional pedagogical project, but should also count on the participation of professionals from different areas in developing the instrument and conducting the process, based on a perspective of formative evaluation, aimed at improving student learning.

Keywords: Higher Education, Educational Assessment, Psychometry, Learning assessment.

INTRODUÇÃO

A importância do movimento constante entre avaliação externa e autoavaliação nas Instituições de Ensino Superior (IES) tornou-se consenso nos diferentes âmbitos da avaliação educacional (Saul, 2002; Depresbiteris, 2005). A Lei do Sistema Nacional de Avaliação da Educação Superior (Sinaes), aprovada em 2004 (Brasil, 2004), soma-se, em suas finalidades, à história de avaliação em muitas IES, a exemplo da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), lócus de desenvolvimento do presente estudo. Nesse contexto, iniciou-se a institucionalização, desde 2006, de um processo único de avaliação de disciplinas dos cursos de graduação da universidade, como elemento da autoavaliação institucional. Um dos aspectos que mais gerou discussões e necessidade de envolvimento da comunidade acadêmica, ao longo da implantação e consolidação desse processo, foi a elaboração dos instrumentos de avaliação.

A qualidade dos instrumentos de medida, conforme destaca Vianna (1976), é fundamental para o desenvolvimento de qualquer atividade científica. No campo da Educação, apesar de a mensuração ser um desafio constante na vida de alunos e professores, o julgamento do desempenho escolar e a atribuição de graus normalmente se dá por meio de instrumentos cuja qualidade, muitas vezes, não foi avaliada. Por outro lado, na área da Psicologia a validação de instrumentos é uma condição necessária para sua credibilidade e difusão. De forma mais específica, a Psicometria é a área que se ocupa desse tema. Cabe salientar, no entanto, que a utilização da Psicometria em procedimentos para validar instrumentos de avaliação de disciplinas não significa assumir exclusivamente o paradigma psicométrico como fundamento dos processos de autoavaliação (Fernandes, 2009).

Dessa forma, a concepção de avaliação que fundamenta o projeto pedagógico da PUCRS pode ser considerada uma avaliação formativa em sentido pleno, ou seja, uma avaliação que assume uma função mais ampla do que a classificação e a certificação, uma vez que se orienta para promover a aprendizagem dos estudantes (Fernandes, 2007). Nesse contexto, a avaliação de disciplinas, compreendida como um processo institucional potencialmente emancipatório é integrante do processo formativo que se desenvolve no ensino de graduação. Para que tal potencialidade se realize, o reconhecimento da qualidade dos instrumentos por professores e alunos é aspecto fundamental para a legitimidade do processo de avaliação, reforçando a importância da validação dos instrumentos.

Com tal enfoque e visando ao aperfeiçoamento do processo de avaliação em sua dimensão formativa, este estudo tem por objetivo apresentar as etapas de desen-

volvimento e validação de um instrumento de avaliação de disciplinas submetido a alunos de graduação a partir da utilização de conceitos extraídos do campo da Psicometria. O estudo está assim organizado: na primeira seção é realizado um breve relato histórico da avaliação do desempenho docente e/ou de disciplinas; na sequência são apresentados o embasamento teórico e as técnicas necessárias à validação do instrumento e à análise da fidedignidade das escalas do mesmo. A seção seguinte apresenta os principais resultados obtidos com base na aplicação das referidas técnicas. Por fim, nas “Considerações Finais” são resumidos os aspectos mais importantes identificados ao longo do estudo.

BREVE RELATO HISTÓRICO

Estudos sobre a validade de mensurar a qualidade instrucional a partir da avaliação de estudantes foram objeto de severas críticas na década de 70 (Greenwald, 1997), visto que havia preocupação de que a avaliação dos estudantes pudesse ser influenciada por um grande número de variáveis a ponto de contaminar os resultados. Assim, variáveis de interesse poderiam estar, de fato, mascarando outros tipos de informação não presentes no instrumento de coleta de dados. Marsh (1984) faz uma listagem de vícios potenciais no processo de avaliação de disciplinas, dentre as quais se destacam: (a) interesse prévio no assunto da disciplina; (b) nota atual e/ou nota esperada; (c) razões por ter se matriculado na disciplina; (d) carga de trabalho e dificuldade na disciplina; e (e) tamanho da turma.

Há, ainda, o original estudo de Naftulin, Ware e Donnelly (1973) no qual os autores mostraram que o estilo pessoal do professor e a sua desenvoltura podem seduzir os estudantes, dando-lhes a ilusão de terem aprendido, mesmo quando falta conteúdo à exposição. Tal influência foi denominada de “efeito Dr. Fox”, em alusão ao ator profissional que conduziu aulas a um grupo de estudantes.

Marsh (1984) ressalta, contudo, que há uma série de estudos mostrando que o efeito de tais variáveis sobre os resultados tende a ser pequeno (McKeachie, 1979; Aleamoni, 1981). O próprio trabalho referenciado como “efeito Dr. Fox” é criticado por ser metodologicamente fraco, cujos resultados obtidos não são passíveis de generalização.

A partir dos anos 80, especialistas em avaliação educacional passaram a considerar medidas quantitativas derivadas da avaliação de estudantes como valiosas fontes de informação. De acordo com Marsh (1984), avaliações provenientes de estudantes podem ser utilizadas com distintas finalidades, tais como: (a) diagnosticar e retroalimentar a instituição sobre a efetividade de seu corpo docente; (b) quantificar

a efetividade do docente com vistas à sua promoção; (c) prover informação aos estudantes buscando orientá-los na seleção de disciplinas e instrutores; e (d) propiciar atividades de pesquisa em Educação. Marsh (1984) reconhece, ainda, que somente a primeira finalidade é praticada pela maior parte das universidades, o mesmo não ocorrendo com as demais.

O uso de dados de avaliação de disciplinas em atividades de pesquisa passou a ser considerado um campo promissor. De fato, muitos estudos foram realizados na década de 80 sobre o tema. A principal conclusão obtida com tais estudos foi a de que a avaliação de disciplinas mostra-se claramente multidimensional, bastante confiável, razoavelmente válida, pouco contaminada por variáveis vistas como potenciais fontes de vício e de extrema importância para professores, alunos e gestores de instituições de ensino.

Apesar de a maior parte dos trabalhos sobre avaliação de professores ter sido publicada entre 1975 e 1990 (Greenwald, 1997), estudos mostrando a validade e a importância da utilização de resultados de avaliação de disciplinas continuaram a ser publicados, embora em menor número. A partir da década de 90, foram conduzidos muitos estudos visando a identificar dimensões subjacentes aos resultados da avaliação de disciplinas, sendo a maioria deles realizada por pesquisadores norte-americanos.

Técnicas estatísticas consagradas em outras áreas do conhecimento e amplamente utilizadas por pesquisadores educacionais norte-americanos ainda não são de uso frequente no Brasil nesse campo do conhecimento. Em nosso país, há poucos trabalhos específicos sobre validação de instrumentos na área de Educação, embora a prática de avaliação do desempenho docente venha sendo adotada pela maioria das instituições de ensino superior, como destacam Nunes e Helfer (2009). Gatti (2004) realizou um levantamento em periódicos da área da Educação, concluindo que poucos empregam metodologias quantitativas. Vianna (2004) faz uma discussão teórica sobre as dimensões envolvidas no processo de avaliação de cursos universitários pelos próprios alunos, na mesma linha de Teixeira e Andrade Jr. (1983). Dentre os estudos quantitativos, destacam-se o de Vieira, Milach e Huppés (2008) que utilizou equações estruturais para modelar a satisfação de alunos em um estudo de caso no curso de Ciências Contábeis e o de Peres-dos-Santos e Laros (2007), que descreve o processo de construção de um instrumento para avaliação da prática pedagógica do professor no ensino superior. Bessa (2007) apresenta um estudo completo sobre validade e metodologia de validação, contudo o foco do estudo são provas educacionais produzidas artesanalmente ou geradas por programas de computador.

AValiação DE DISCIPLINAS: UM PROBLEMA MULTIDIMENSIONAL

Entende-se por dimensão subjacente ou constructo um conceito teórico que não existe no sentido literal ou físico. Apesar disso, ele pode ser medido por meio de um conjunto de itens, também chamados de variáveis manifestas, que fornecem evidências desses constructos (Hair et al., 1998). Marsh (1984) considera a avaliação do desempenho docente um fenômeno inequivocamente multidimensional.

Koch (1981) investigou dimensões subjacentes à avaliação do desempenho de professores em uma amostra de 663 estudantes da Universidade do Texas, utilizando as técnicas estatísticas de Análise Fatorial e Escalonamento Multidimensional para detectar quatro fatores subjacentes em um conjunto de 34 itens: 1) aspectos relacionados à disciplina, 2) atributos e estilo do professor, 3) relacionamento entre professor e aluno, e 4) relevância da disciplina para o curso.

Teixeira e Andrade Jr. (1983) abordaram a avaliação do desempenho do corpo docente visualizando-a num contexto mais abrangente e inserida num sistema de avaliação do ensino-aprendizagem. Eles destacaram quatro dimensões que compõem o ensino-aprendizagem: 1) aluno, 2) professor, 3) objetivos, e 4) matéria e método. Peres-dos-Santos e Laros (2007) identificaram dois fatores – “relacionamento entre professor e aluno” e “prática de ensino” – em um conjunto de 16 itens que compunha um instrumento de avaliação da prática pedagógica do professor no ensino superior.

Petchers e Chow (1988) utilizaram análise fatorial exploratória para encontrar três dimensões na avaliação do desempenho docente: 1) mérito do instrutor, 2) processo de avaliação ou *feedback*, e 3) importância da disciplina. Jackson et al. (1999), por sua vez, identificaram dois fatores gerais ao aplicar o instrumento SPTE (*Students' Perceptions of Teaching Effectiveness*) em uma amostra de 7 mil estudantes: 1) qualidade geral do ensino, e 2) demandas do curso. No mesmo estudo, os autores compararam os resultados por eles encontrados com outros trabalhos similares, identificando substanciais similaridades.

No presente estudo, o número de itens que compõe o instrumento não poderia ser excessivo (ver detalhes na seção Resultados, subseção Validade de Conteúdo). A proposição do instrumento de avaliação de disciplinas considerou as restrições relativas à sua extensão, os estudos encontrados na literatura e as reuniões com especialistas para validação de conteúdo. O instrumento foi concebido de forma a contemplar três dimensões: 1) aspectos gerais da disciplina, 2) ação de ensino do professor, e 3) autoavaliação do aluno. Tais dimensões apresentam-se em harmonia com os principais estudos da área e atendem às necessidades institucionais da PUCRS.

MÉTODO

O método proposto neste estudo baseia-se em técnicas de fidedignidade e de validação de instrumentos utilizadas no campo da Psicometria (Fachel; Camey, 2000). Conforme relatado na seção anterior, o uso de tais técnicas é justificado pelo fato de que a avaliação educacional traz consigo um conjunto de dimensões subjacentes que não podem ser diretamente mensuradas (Sousa, 2000).

Validade e Fidedignidade

A qualidade do instrumento proposto foi avaliada por meio de técnicas de validação de conteúdo, de critério, de constructo e de estabilidade temporal, além de coeficientes de fidedignidade.

A validade de um instrumento ou teste refere-se ao fato de ele realmente medir aquilo a que se propõe (Lindeman, 1974). De acordo com Fachel e Camey (2000) o conceito de validade é frequentemente definido por meio da seguinte pergunta: você está medindo o que pensa que está medindo? Vianna (1976) distingue quatro tipos de validade (de conteúdo, concorrente, preditiva e de constructo). Todavia, no presente estudo foi adotada a classificação de Anastasi (1988) que sugere que há apenas três categorias, com a seguinte nomenclatura: validade de conteúdo, de critério e de constructo.

A validade de conteúdo não é determinada por medidas estatísticas, visto que ela resulta do julgamento de juízes ou pessoas reconhecidas na área do estudo, que analisam a representatividade dos itens em relação aos conceitos que se pretende medir. Já a validade de critério está relacionada à capacidade de o instrumento ou teste funcionar como um preditor, presente ou futuro, de outra variável operacionalmente independente chamada critério (Fachel; Camey, 2000). Quando os resultados da aplicação de um instrumento são comparados com os de outro instrumento similar, obtidos na mesma época, procede-se à validade concorrente. No caso de resultados presentes serem utilizados na previsão de um diagnóstico futuro, o procedimento é denominado validade preditiva. Tanto esta quanto a validade concorrente são subdivisões da validade de critério. Por fim, a validade de constructo está relacionada ao grau com que um instrumento mede o que lhe foi designado medir. Ela baseia-se no fato de que os conceitos a serem medidos são variáveis latentes não-observáveis. Para identificação dos itens que compõem cada dimensão subjacente de um instrumento, a técnica estatística multivariada amplamente utilizada é a Análise Fatorial.

Além da validade, outra qualidade necessária a um instrumento de medida é a sua fidedignidade, ou seja, as medidas dele derivadas devem ser consistentes, pre-

cisas e estáveis. Isso quer dizer que um instrumento ou teste fidedigno levaria aos mesmos resultados se aplicado ao mesmo grupo por repetidas vezes (Vianna, 1976). Existem vários coeficientes para se estimar a fidedignidade, baseados em dois diferentes conceitos: consistência interna e estabilidade temporal. Neste estudo, para o primeiro conceito foi utilizado o coeficiente de fidedignidade *Alpha de Cronbach*; para a estabilidade temporal foi usado o coeficiente de variação.

O quadro 1 apresenta o esquema de validação e as técnicas ou procedimentos utilizados em cada um dos aspectos considerados.

Quadro 1 – Esquema de validação do instrumento de avaliação de disciplinas

Característica	Tipo	Procedimento/Técnica
Validade	Conteúdo	Reuniões com especialistas
	Construto	Análise Fatorial Confirmatória
	Critério	Regressão Linear Múltipla
Fidedignidade	Consistência interna	Alpha de Cronbach e Variância extraída
	Estabilidade temporal	Coefficiente de Variação

Os itens a seguir apresentam separadamente as técnicas utilizadas no presente estudo.

Reuniões com especialistas

Considerando-se que a validade de conteúdo não é determinada por meio de medidas quantitativas faz-se necessário reunir uma comissão multidisciplinar de especialistas no tema em estudo com o objetivo de analisar o alinhamento do instrumento aos pressupostos teóricos. Além de profissionais da área de Educação é desejável contar com especialistas nas áreas de Letras, Psicologia e Estatística. A discussão entre profissionais oriundos de diferentes áreas é uma etapa fundamental no processo de definição do instrumento, visto que todas as medidas estatísticas utilizadas nas demais etapas de validação dependem dessa definição.

Análise Fatorial Confirmatória

A Análise Fatorial (AF) é uma técnica estatística multivariada amplamente utilizada na área das ciências do comportamento e, em especial, na validação de construtos. Ela tem por objetivo descrever a estrutura de correlação de um conjunto de

variáveis observadas na forma de um pequeno número de novas variáveis subjacentes, conhecidas como fatores. A premissa que rege a AF é a de que variáveis que tenham um significado comum entre si apresentem alto nível de correlação. A técnica de AF, que pode ser de caráter exploratório ou confirmatório, teve início no campo da Psicometria em estudos que tinham por objetivo definir e medir “inteligência”. Segundo Johnson e Wichern (2001), a AF exploratória tem gerado controvérsias e turbulências desde sua criação, em razão do alto índice de subjetividade associado à interpretação de seus resultados.

A partir da década de 80, a validação de construto começou a ser realizada por meio da Análise Fatorial Confirmatória (AFC), o que reduziu a subjetividade presente na análise exploratória. No caso da AFC, a composição das dimensões subjacentes aos itens de um instrumento é definida pelo próprio pesquisador e as cargas fatoriais (peso das variáveis em cada dimensão) são testadas individualmente.

A AFC possibilita que o pesquisador defina a estrutura fatorial *a priori*, podendo analisá-la por meio de testes de significância específicos (Jöreskog; Sörbom, 1993). Neste estudo, a AFC foi utilizada para a validação estatística da estrutura fatorial prevista no instrumento de avaliação de disciplinas.

As estimativas das cargas fatoriais foram geradas pelo *software* LISREL[®], assim como as medidas de adequabilidade amostral. Tanto os testes das cargas fatoriais como as medidas de ajustamento permitem verificar se os dados observados apresentam aderência ao modelo fatorial idealizado teoricamente. As quatro medidas de adequabilidade utilizadas foram:

- Raiz quadrada média do erro de aproximação (RMSEA) é uma medida da falta de ajuste entre a matriz de covariância empírica e a estimada pelo modelo fatorial. Um valor alto de RMSEA indica que as duas matrizes são substancialmente diferentes. Para um bom ajuste, uma medida RMSEA < 0,08 é considerada adequada (Chen et al., 2008). Outros autores (McCallum; Browne; Sugawara, 1996) sugerem que os valores 0,01, 0,05 e 0,08 indicam ajustes excelentes, bons e medíocres, respectivamente.
- *Goodness-of-fit index* (GFI) faz uma comparação entre o modelo final estimado e o modelo inicial. Valores próximos de 1 indicam bom ajustamento.
- *Normal Fit Index* (NFI) é uma medida que varia de 0 a 1, sendo que 1 indica ajustamento perfeito. Apesar de não existir base estatística, valores iguais ou superiores a 0,90 são desejáveis (Hair et al., 1998).

- *Nonnormed Fit Index* (NNFI) é uma medida similar ao NFI que corrige uma tendência deste a subestimar o ajuste. Valores superiores a 0,90 são recomendados.
- A medida da estatística *qui-quadrado*, dividida pelos seus respectivos graus de liberdade, não foi usada, em razão do grande tamanho amostral, o que causa uma inflação dessa estatística.

Análise de Regressão Múltipla

A análise de regressão múltipla é uma técnica estatística multivariada para analisar o relacionamento entre um conjunto de variáveis independentes (X_i) com uma única variável dependente (Y). O modelo de regressão mais comumente utilizado é o linear, com a seguinte forma:

$$Y = \alpha + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$$

onde os coeficientes β_i indicam o peso da variável independente X_i sobre a variável dependente Y (Hair et al., 1998). O termo α é uma constante para adequação de escala. Em termos práticos, quando todas as variáveis independentes assumem o valor 0, tem-se $Y=\alpha$. O termo ε , ou resíduo, representa a diferença entre o valor efetivamente observado e o valor estimado.

Para avaliar o impacto das variáveis independentes X_i sobre a variável dependente Y , é comum utilizar os coeficientes β padronizados, pois os mesmos possibilitam uma comparação direta entre si, visto que não sofrem efeitos de escala. A habilidade preditiva do modelo é aferida por meio do coeficiente de determinação (R^2), o qual indica o percentual de variações de Y que pode ser explicado pelo conjunto de variáveis independentes. Além do R^2 , há também testes de significância específicos para os coeficientes de cada variável independente X_i e para o modelo completo, que resultam significativos quando o valor de probabilidade associado (p) for inferior ao nível de significância adotado.

Coefficiente *Alpha* de Cronbach e variância extraída

Proposto originalmente por Cronbach (1951), o coeficiente *Alpha* é uma medida utilizada em instrumentos psicométricos para avaliar a consistência interna de dimensões ou construtos. A medida é baseada na correlação dos itens que compõem

cada dimensão. De acordo com Hair et al. (1998), valores acima de 0,70 indicam consistência interna satisfatória.

Esses autores sugerem, ainda, que o *Alpha de Cronbach* não seja utilizado como único indicador de fidedignidade. A variância extraída dos fatores, baseada na magnitude das cargas fatoriais, é sugerida por Hair et al. (1998) como alternativas melhores. Valores acima de 50% são considerados aceitáveis.

Estabilidade temporal

Com o objetivo de analisar a estabilidade temporal dos resultados da avaliação de disciplinas, as cargas fatoriais e as medidas de ajustamento relativas a cada um dos semestres 2008/1, 2008/2 e 2009/1 foram estimadas separadamente. A comparação entre elas foi realizada por meio do coeficiente de variação (CV), que consiste na razão entre o desvio-padrão e a média do conjunto de dados. A estabilidade temporal é identificada por valores pequenos do CV, indicando que as estimativas nos diferentes semestres tendem a orbitar em torno de um mesmo valor. Esta pode ser considerada uma maneira indireta de verificar a estabilidade temporal dos resultados de validação do instrumento.

Coleta de dados e amostra

Os dados utilizados neste estudo correspondem a resultados de avaliação de disciplinas da PUCRS nos semestres 2008/1, 2008/2 e 2009/1, com um total de 151.220 instrumentos preenchidos por alunos. Tratando-se de avaliação de disciplinas, um mesmo aluno pode vir a preencher mais de um instrumento. Em média, cada participante respondeu a 4,7 instrumentos. A tabela 1 apresenta a caracterização do número de alunos respondentes em cada edição da avaliação.

Tabela 1 – Número absoluto e relativo de alunos participantes e instrumentos respondidos no processo de avaliação de disciplinas da PUCRS (2008/1, 2008/2, 2009/1)*

Edição	2008/1		2008/2		2009/1		Total
	n	%	n	%	n	%	
Alunos respondentes	11.551	44,4	11.118	44,3	9.455	37,4	32.124
Instrumentos respondidos	54.976	39,5	51.804	38,8	44.440	32,7	151.220
Razão instrumentos/aluno	4,8	-	4,7	-	4,7	-	4,7

Fonte: PUCRS.

* Os percentuais foram calculados em relação ao máximo possível de alunos participantes ou de instrumentos preenchidos.

O sistema de avaliação de disciplinas da PUCRS foi disponibilizado ao aluno durante as últimas seis semanas dos períodos letivos, de forma que o mesmo tivesse uma visão geral da disciplina. O processo de amostragem caracterizou-se como não-probabilístico, pois a composição da amostra deu-se a partir da conveniência do respondente em participar da pesquisa (Malhotra, 2008). Nesse sentido, a amostra da avaliação de disciplinas resultou do interesse do aluno por acessar o sistema de pesquisas da Universidade e por responder ao questionário. A escala de respostas utilizada foi ordinal de cinco pontos, popularmente conhecida como tipo *Likert*, onde 1 significava “Muito Insatisfeito” e 5 “Muito Satisfeito”.

RESULTADOS

Nesta seção, os resultados são apresentados de acordo com a estrutura de validação e de fidedignidade mostrada anteriormente no quadro 1. Cada um dos tipos de validade e de fidedignidade do instrumento de avaliação de disciplinas é discutido separadamente.

Validade de conteúdo

O instrumento de avaliação de disciplinas, utilizado na PUCRS até o 2º semestre de 2007, continha um total de 28 questões. Como o instrumento devia ser preenchido separadamente por disciplina, a extensão do mesmo revelou-se um problema, uma vez que um aluno matriculado em cinco disciplinas, por exemplo, necessitaria responder a 140 questões. No início de 2008, a comissão de especialistas responsável pela avaliação institucional na Universidade definiu que um novo instrumento deveria ser elaborado, considerando-se as mesmas dimensões do anterior, mas com menor número de questões.

O novo instrumento foi submetido a um pré-teste com grupos de alunos, antes da primeira aplicação em 2008/1. O conteúdo foi validado por gestores, docentes e especialistas da área de Educação e avaliação, bem como por linguistas, com a preocupação de alinhar o instrumento aos pressupostos teóricos que fundamentam a avaliação de disciplinas na PUCRS.

A partir de 2008/1, portanto, o instrumento de avaliação de disciplinas respondido pelos alunos dos cursos de graduação passou a ter 13 questões, organizadas em três blocos, além de uma questão de avaliação geral da disciplina e de um espaço para comentários, conforme a descrição que segue.

1º bloco – *Aspectos gerais da disciplina*: o aluno avalia a relevância da disciplina para o desenvolvimento profissional e pessoal, a sua relação com as demais do curso, a organização do plano da mesma e a indicação de textos e de outros materiais de estudo.

2º bloco – *Ação de ensino do professor*: o aluno indica o seu grau de satisfação em relação à clareza do professor no desenvolvimento do conteúdo, aos procedimentos didáticos, à reflexão sobre os resultados das atividades de avaliação dessa disciplina, à contribuição para um ambiente favorável à aprendizagem e ao aproveitamento do tempo da aula.

3º bloco – *Autoavaliação do aluno*: o aluno indica o seu grau de satisfação em relação ao seu desempenho quanto à utilização dos materiais de estudo indicados, à sua motivação para aprender mais, à sua contribuição, assim como de sua turma, para um ambiente favorável à aprendizagem.

Avaliação geral: na perspectiva de uma síntese da avaliação, a última questão refere-se à avaliação geral da disciplina.

Espaço para comentários: nesse espaço, o aluno pode expressar-se livremente acerca da disciplina. Os comentários são organizados em um banco de dados qualitativo e analisados, propiciando uma compreensão mais ampliada sobre a percepção que o discente tem em relação à disciplina.

Validade de construto – Análise Fatorial Confirmatória

A Análise Fatorial Confirmatória (AFC) considerou os 151.220 instrumentos preenchidos nos três semestres letivos. A matriz de correlação entre as 13 questões do instrumento consistiu no principal dado de entrada da AFC. A tabela 2 mostra a estrutura do modelo fatorial idealizado teoricamente, onde $\lambda_{i,j}$ representa a carga fatorial da questão i no fator j . A tabela 3 apresenta as estimativas das cargas fatoriais obtidas por meio do *software* LISREL® pelo método da máxima verossimilhança.

Tabela 2 – Estrutura do modelo fatorial idealizada teoricamente

	Fator 1 – Aspectos relacionados à disciplina	Fator 2 – Ação de ensino do professor	Fator 3 – Autoavaliação dos alunos
Q1 Relevância desta disciplina para o desenvolvimento profissional/pessoal	$\lambda_{1,1}$		
Q2 Relação desta disciplina com as demais do curso	$\lambda_{2,1}$		
Q3 Organização do plano da disciplina	$\lambda_{3,1}$		
Q4 Indicação de textos e de outros materiais de estudo	$\lambda_{4,1}$		
Q5 Clareza no desenvolvimento do conteúdo		$\lambda_{5,2}$	
Q6 Procedimentos didáticos		$\lambda_{6,2}$	
Q7 Reflexão sobre os resultados das atividades de avaliação desta disciplina		$\lambda_{7,2}$	
Q8 Contribuição para um ambiente favorável à aprendizagem		$\lambda_{8,2}$	
Q9 Aproveitamento do tempo da aula		$\lambda_{9,2}$	
Q10 Utilização dos materiais de estudo indicados			$\lambda_{10,3}$
Q11 Motivação para aprender mais			$\lambda_{11,3}$
Q12 Contribuição como aluno para um ambiente favorável à aprendizagem			$\lambda_{12,3}$
Q13 Contribuição da turma para um ambiente favorável à aprendizagem			$\lambda_{13,3}$

Tabela 3 – Cargas fatoriais das questões em cada fator estimadas por máxima verossimilhança no software LISREL®

	Fator 1 – Aspectos relacionados à disciplina	Fator 2 – Ação de ensino do professor	Fator 3 – Autoavaliação dos alunos
Q1 Relevância desta disciplina para o desenvolvimento profissional/pessoal	0,705**		
Q2 Relação desta disciplina com as demais do curso	0,747**		
Q3 Organização do plano da disciplina	0,725**		
Q4 Indicação de textos e de outros materiais de estudo	0,754**		
Q5 Clareza no desenvolvimento do conteúdo		0,830**	
Q6 Procedimentos didáticos		0,804**	
Q7 Reflexão sobre os resultados das atividades de avaliação desta disciplina		0,815**	
Q8 Contribuição para um ambiente favorável à aprendizagem		0,877**	
Q9 Aproveitamento do tempo da aula		0,772**	
Q10 Utilização dos materiais de estudo indicados			0,635**
Q11 Motivação para aprender mais			0,800**
Q12 Contribuição como aluno para um ambiente favorável à aprendizagem			0,802**
Q13 Contribuição da turma para um ambiente favorável à aprendizagem			0,791**

** Cargas fatoriais significativas pelo teste de Wald ao nível de 1% ($p < 0,01$).

Todas as cargas fatoriais revelaram-se significativas, indicando que a alocação das questões aos devidos fatores mostrou-se adequada. À exceção de Q10, as demais questões apresentaram carga fatorial superior a 0,70, o que reforça a adequabilidade do modelo, visto que elas variam no intervalo de [-1;+1].

A tabela 4 apresenta os valores das quatro medidas de ajustamento descritas na seção anterior, sendo que todas estão dentro dos limiares encontrados na literatura.

Tabela 4 – Medidas de ajustamento do modelo fatorial

Medidas	RMSEA	GFI	NFI	NNFI
Valores	0,022	0,995	0,942	0,928

Com base nos resultados da AFC, pode-se inferir que os construtos do instrumento de avaliação de disciplinas foram estatisticamente validados.

Validade de critério – Análise de Regressão Múltipla

Para haver validade de critério, os escores derivados do instrumento ou do teste proposto deverão funcionar como preditores de outra variável, operacionalmente independente, chamada de critério. Neste estudo, a última questão, de avaliação geral da disciplina, foi considerada como critério e os escores dos três fatores formaram o conjunto de variáveis predictoras. Uma análise de regressão múltipla permitiu inferir sobre a habilidade preditiva do instrumento em relação à avaliação geral da disciplina. A tabela 5 apresenta os resultados.

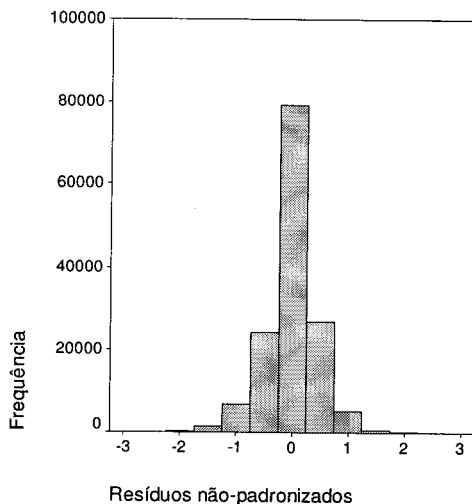
Tabela 5 – Análise de regressão múltipla da avaliação geral da disciplina em função dos três fatores extraídos e das medidas de ajustamento do modelo

Parâmetros	β	Erro-padrão do β	β padronizado	Valor de p
Constante	-0,2742	0,0058	-	< 0,001
Fator 1	0,0915	0,0008	0,2270	< 0,001
Fator 2	0,1525	0,0005	0,6156	< 0,001
Fator 3	0,0532	0,0007	0,1304	< 0,001

Coefficiente de determinação: $R^2 = 83,1\%$; Significância do modelo: $p < 0,001$

O modelo de regressão foi altamente significativo do ponto de vista estatístico, uma vez que 83,1% das variações na avaliação geral da disciplina podem ser atribuídas aos escores obtidos nos três fatores. Como todos os coeficientes apresentam sinais positivos, a relação entre os fatores e a avaliação geral da disciplina é direta, ou seja, um aumento nos escores dos fatores implica incremento da avaliação geral. A figura 1 mostra que as diferenças entre os valores observados e estimados pelo modelo de regressão (resíduos) apresentam pequena variação em torno de zero, o que reforça a habilidade preditiva do modelo.

Figura 1 – Histograma dos resíduos do modelo de regressão



O ponto ideal do coeficiente de determinação (R^2) em termos de validade de critério seria 100%, entretanto essa cifra é utópica em se tratando de estudos empíricos. Assim, um $R^2 = 83,1\%$ é considerado amplamente satisfatório, apontando para a validade do instrumento proposto.

Não foi possível realizar o procedimento de validade concorrente, em razão do necessário sigilo das informações individualizadas. Como todas as respostas são armazenadas sem a identificação do respondente, não se poderia aplicar outros instrumentos numa amostra de alunos para proceder a cálculos de correlação. Pelo mesmo motivo, também não se pode realizar procedimentos de teste e reteste.

Análise de fidedignidade: *Alpha de Cronbach* e variância extraída

As medidas de fidedignidade utilizadas neste estudo são apresentadas na tabela 6 separadamente para cada fator, sendo que todas elas excederam os limiares sugeridos na literatura. Como o instrumento mostrou-se válido, pode-se concluir que ele efetivamente mede aquilo a que se propõe de maneira consistente.

Tabela 6 – Medidas de fidedignidade dos fatores: *Alpha de Cronbach* e variância extraída

Fatores	Número de itens	Alpha de Cronbach	Variância extraída
Fator 1 – Aspectos relacionados à disciplina	4	0,874	54%
Fator 2 – Ação de ensino do professor	5	0,948	67%
Fator 3 – Autoavaliação dos alunos	4	0,875	58%

Estabilidade temporal

A estabilidade temporal do instrumento de avaliação foi verificada por meio do coeficiente de variação (CV), conforme descrito na seção “Método”. Valores de CV baixos indicam que as qualidades psicométricas do instrumento mantêm-se constantes, independentemente do período de coleta de dados ou do conjunto de respondentes.

A tabela 7 apresenta as cargas fatoriais das três AFC realizadas separadamente por semestre letivo. Todos os 13 itens que compõem o instrumento revelaram cargas fatoriais estáveis, com variações que não ultrapassaram 2,5%.

Tabela 7 – Cargas fatoriais estimadas separadamente por semestre e coeficientes de variação

Fator	Questão	Ano / Semestre			Coefficiente de Variação %
		2008/1	2008/2	2009/1	
Fator 1 – Aspectos relacionados à disciplina	Q1	0,726	0,705	0,695	2,2
	Q2	0,762	0,748	0,738	1,6
	Q3	0,738	0,726	0,718	1,4
	Q4	0,764	0,756	0,747	1,1
Fator 2 – Ação de ensino do professor	Q5	0,834	0,830	0,828	0,4
	Q6	0,812	0,805	0,800	0,7
	Q7	0,822	0,815	0,812	0,6
	Q8	0,883	0,878	0,873	0,6
	Q9	0,785	0,773	0,763	1,4
Fator 3 – Autoavaliação dos alunos	Q10	0,653	0,641	0,621	2,5
	Q11	0,808	0,807	0,792	1,1
	Q12	0,809	0,806	0,795	0,9
	Q13	0,798	0,795	0,783	1,0

As tabelas 8 e 9 apresentam as medidas de ajustamento para o modelo fatorial e as medidas de fidedignidade nos três semestres, todas acompanhadas pelos respectivos coeficientes de variação (CV).

Tabela 8 – Medidas de ajustamento do modelo fatorial por semestre e CV

Medidas	RMSEA	GFI	NFI	NNFI
2008/1	0,020	0,996	0,949	0,938
2008/2	0,021	0,995	0,943	0,931
2009/1	0,023	0,995	0,938	0,925
Coeficiente de Variação	7,2%	0,1%	0,6%	0,7%

Tabela 9 – Fidedignidade e variância extraída por semestre e CV

Medidas	Ano/Semestre	Fator 1 – Aspectos relacionados à disciplina	Fator 2 – Ação de ensino do professor	Fator 3 – Autoavaliação dos alunos
Fidedignidade	2008/1	0,835	0,916	0,852
	2008/2	0,824	0,912	0,849
	2009/1	0,816	0,909	0,837
	CV	1,16%	0,38%	0,94%
Variância Extraída	2008/1	55,9%	68,5%	59,3%
	2008/2	53,9%	67,4%	58,6%
	2009/1	52,5%	66,7%	56,5%
	CV	3,16%	1,34%	2,50%

Das quatro medidas de ajustamento do modelo fatorial, três delas apresentaram CV abaixo de 1%. A exceção foi a RMSEA que apresentou uma variação de 7,2%, em razão da pequena magnitude dos valores. Na realidade, a variação absoluta foi de apenas 0,003 pontos.

Em relação às medidas de fidedignidade do instrumento, todas mostraram coeficientes de variação iguais ou inferiores a 2,5%, à exceção da variância extraída do Fator 1, cujo valor foi ligeiramente superior (3,16%). De forma geral, com base

nos resultados dos CV, pode-se considerar que o instrumento de avaliação proposto apresenta estabilidade temporal, o que reforça sua consistência.

CONSIDERAÇÕES FINAIS

Este artigo apresentou as etapas do processo de desenvolvimento e de validação de um instrumento de avaliação de disciplinas de graduação, utilizando técnicas oriundas do campo da Psicometria. Buscou-se, também, contextualizar historicamente o problema proposto, trazendo estudos seminais realizados na área, bem como relatar como a avaliação de disciplinas insere-se no processo de autoavaliação da PUCRS.

Quatro aspectos devem ser destacados nestas considerações finais:

1) Relevância da avaliação de disciplinas: O conjunto de informações constituído pelas avaliações de disciplinas de uma IES é uma importante fonte de informação que merece ser explorada. Dentre as quatro finalidades listadas por Marsh (1984) – (a) diagnosticar e retroalimentar a instituição sobre a efetividade de seu corpo docente; (b) quantificar a efetividade do docente com vistas à sua promoção; (c) prover informação aos estudantes buscando orientá-los na seleção de disciplinas e instrutores; e (d) propiciar atividades de pesquisa em Educação – pode-se afirmar que a primeira é a finalidade mais frequente dentre as IES brasileiras. O caráter sigiloso e eletivo das informações prestadas pelo aluno – estratégia utilizada pela maior parte das IES para aumentar a confiabilidade nos resultados – faz com que as finalidades (b) e (c) sejam práticas pouco adotadas no cenário educacional brasileiro. A opção pelo sigilo e pela eletividade da participação não prejudica a utilização dos resultados para diagnóstico e retroalimentação da instituição, tampouco impossibilita sua utilização para fins de pesquisa, última finalidade listada por Marsh (1984).

A apresentação de medidas estatísticas descritivas relativas às dimensões presentes no instrumento proposto permite ao corpo docente vislumbrar, de forma sintética, a avaliação da ação ensino-aprendizagem sob a visão do aluno. Os resultados de avaliação propiciam uma reflexão entre professores e gestores, contribuindo para aprimorar o projeto pedagógico institucional.

2) O instrumento proposto e suas características psicométricas: A validação do instrumento revelou-se uma importante iniciativa para o aperfeiçoamento do processo de avaliação de disciplinas. Neste estudo, o instrumento demonstrou

possuir todas as qualidades psicométricas desejáveis, sendo, portanto, considerado válido e fidedigno. A consistência do tamanho amostral (32.124 alunos e 151.220 instrumentos) e a estabilidade dos resultados obtidos nos três períodos letivos reforçam a credibilidade do processo de avaliação.

Deve-se ressaltar, contudo, que apesar de o instrumento proposto ter sido considerado válido e fidedigno, ele ainda pode vir a ser aprimorado em estudos futuros, visto que não existem instrumentos definitivos e perfeitos, especialmente em se tratando de avaliação educacional.

3) Avaliação formativa: Compreender a complexidade das interações inerentes ao processo avaliativo, que inclui a adesão ao mesmo, o preenchimento do instrumento, o acesso e a utilização dos resultados, é um campo de investigação que, de modo complementar, emerge no contexto de trabalhos dessa natureza. Muitos são os desafios desse processo em termos de reconhecimento, por parte da comunidade acadêmica, de sua dimensão formativa. Resistências, incompreensões e dificuldades de aceitação, entre outros, são aspectos inerentes à formação humana e que se apresentam nesse processo.

Nesse sentido, a coerência entre a dimensão formativa esperada e as práticas efetivamente exercidas na comunidade acadêmica requer um trabalho de acompanhamento e de formação permanente, no qual a reflexão sobre os instrumentos passa a oferecer uma contribuição fundamental.

4) Métodos quantitativos na área da Educação: O estudo mostra claramente que a área da Educação é um terreno fértil para o desenvolvimento da pesquisa quantitativa e, em especial, próprio para a aplicação de técnicas estatísticas multivariadas. Há que se desmitificar o dilema entre os paradigmas qualitativo e quantitativo, de forma que eles não sejam vistos como contraditórios, mas, sim, como complementares. Ambos apresentam vantagens e limitações (Fernandes, 1991).

Por fim, há que se destacar Gatti (2004) quando afirma que métodos quantitativos em educação, contextualizados por teoria sólida e metodologia cuidadosa, trazem subsídios concretos para a compreensão de fenômenos educacionais.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALEAMONI, L. M. Student ratings of instruction. In: MILLMAN, J. (Ed.). *Handbook of teacher evaluation*, Beverly Hills, CA: Sage, 1981. p. 110-145.
- ANASTASI, A. *Psychological testing*. 6.ed. New York: MacMillan, 1988.
- BESSA, N. M. Validade: o conceito, a pesquisa, os problemas de provas geradas pelo computador. *Estudos em Avaliação Educacional*, v. 18, n. 37, p. 115-155, maio/ago. 2007.
- BRASIL. Lei n.º 10.861, de 14 de abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior - Sinaes e dá outras providências. *Diário Oficial [da] República Federativa do Brasil*, Brasília, DF, 15 abr. 2004
- CHEN, F. et al. An Empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, v. 36, n. 4, p. 462-494, 2008.
- CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, n. 16, v. 3, p. 297-334, 1951.
- DEPRESBITERIS, L. Autoavaliação das instituições de ensino superior. *Estudos em Avaliação Educacional*, São Paulo, v. 16, n. 32, p. 7-24, jan./abr. 2005.
- FACHEL, J. M. G.; CAMEY, S. A. Avaliação Psicométrica: a qualidade das medidas e o entendimento dos dados. In: CUNHA, J. A. (Org.). *Psicodiagnóstico*. 5.ed. Porto Alegre, RS: Artes Médicas, 2000, p. 158-170. 5 v.
- FERNANDES, D. Notas sobre os paradigmas da investigação em educação. *Noesis*, Lisboa, v. 18, p. 64-66, 1991.
- _____. A Avaliação das aprendizagens no sistema educativo português. *Educação e Pesquisa*, São Paulo, v. 33, n. 3, p. 581-600, set./dez. 2007.
- _____. *Avaliar para aprender: fundamentos, práticas e políticas*. São Paulo: Unesp, 2009.
- GATTI, B. A. Estudos quantitativos em educação. *Educação e Pesquisa*, São Paulo, v. 30, n. 1, p. 11-30, jan./abr. 2004.
- GREENWALD, A. G. Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, v. 52, n. 11, p. 1182-1186, 1997.
- HAIR, J. F. et al. *Multivariate data analysis*, 5. ed. Englewood Cliffs: Prentice-Hall, 1998.
- JACKSON, D.L. et al. The dimensions of students' perceptions of teaching effectiveness. *Educational and Psychological Measurement*, v. 59, n. 4, p. 580-596, 1999.
- JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. 5.ed. Englewood Cliffs: Prentice-Hall, 2001.
- JÖRESKOG, K. G.; SÖRBOM, D. *LISREL® 8: structural equation modeling with the SIMPLIS™ command language*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers, 1993.
- KOCH, W. Dimensions underlying student ratings of instruction. In: ANNUAL MEETING OF THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, 65., 1981, Los Angeles. *Proceedings...*, Washington D.C.: ERIC, 1981. Disponível em: <<http://www.eric.ed.gov/PDFS/ED204395.pdf>>. Acesso em: 23 ago. 2010.
- LINDEMAN, R. H. *Medidas educacionais*. Porto Alegre: Globo, 1974.
- MALHOTRA, N. K. *Basic marketing research*. 3.ed. Upper Saddle River: Prentice Hall, 2008.
- MARSH, H. W. Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology*, v. 76, n. 5, p. 707-754, 1984.
- MCCALLUM, R. C.; BROWNE, M. W.; SUGAWARA, H. M. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, n. 1, p. 130-149, 1996.
- McKEACHIE, W. J. Student ratings of faculty: a reprise. *Academe*, p. 384-397, 1979.
- NAFTULIN, D. H.; WARE, J. E.; DONNELLY, F. A. The Doctor fox lecture: a paradigm of educational seduction. *Journal of Medical*

Education, v. 48, p. 630-635, 1973.

NUNES, A. K.; HELFER, C. L. L. Diagnóstico do desempenho na docência da graduação da UNISC. *Avaliação*, Campinas, v. 14, n. 1, p. 169-183, mar. 2009.

PERES-DOS-SANTOS, L. F. B.; LAROS, J. A. Avaliação da prática pedagógica do professor de ensino superior. *Estudos em Avaliação Educacional*, São Paulo, v. 18, n. 36, p. 75-95, jan./abr. 2007.

PETCHERS, M. K.; CHOW, J. Interpreting students' course evaluations: a look at the underlying conceptual dimensions. *Journal of Teaching in Social Work*, v. 2, n. 2, p. 51-61, 1988.

SAUL, A. M. A Sistemática de autoavaliação do Programa de Pós-Graduação em Educação (currículo) da PUC/SP. *Estudos em Avaliação Educacional*, n. 26, p. 97-109, jul./dez. 2002.

SOUSA, C. P. Dimensões da avaliação educacional. *Estudos em Avaliação Educacional*, n. 22, p. 101-118, jul./dez. 2000.

TEIXEIRA, G. W.; ANDRADE JR., M. N. Avaliação do processo de ensino-aprendizagem e suas dimensões. *Revista de Administração*, v. 18, n. 4, p. 69-75, 1983.

VIANNA, H. M. *Testes em educação*. 2. ed. São Paulo: Ibrasa, 1976.

_____. Avaliação de cursos pelos alunos: considerações. *Estudos em Avaliação Educacional*, n. 29, p. 137-147, jan./jun. 2004.

VIEIRA, K. M.; MILACH, F. T.; HUPPES, D. Equações estruturais aplicadas à satisfação dos alunos: um estudo no curso de ciências contábeis da Universidade Federal de Santa Maria. *Revista Contabilidade e Finanças*, v. 19, n. 48, p. 65-76, 2008.

Recebido em: dezembro 2010

Aprovado para publicação em: fevereiro 2011