

Validade de um Instrumento de Avaliação de Desempenho de Docentes do Ensino Superior*

RICARDO PRIMI** & ILMARA FÁTIMA DE MORAES***

Recebido: 05/06/05

Aprovado: 08/08/05

** Professor do Programa de Pós Graduação Stricto Sensu em Psicologia, Universidade São Francisco (USF). E.mail: rprimi@uol.com.br

*** Fundação Municipal de Ensino Superior de Bragança Paulista, Doutoranda em Avaliação Psicológica, Universidade São Francisco (USF).

Resumo: Investiga-se validade e precisão de um questionário de avaliação docente, com 15 itens de 5 pontos, utilizado em uma universidade paulista, respondido voluntariamente por estudantes, via Internet, em "ambiente seguro". Foram 3689 estudantes avaliando 568 professores de 36 cursos. Foi realizada análise fatorial, análise de perfis e estudo correlacional da nota escolar com avaliação docente. A precisão evidenciou excelente consistência interna dos itens ($\alpha=0,965$). Os itens agruparam-se em um fator, o que foi confirmado pela análise de perfis. Apesar da natureza multidimensional desse construto os alunos parecem formar uma opinião global sobre o desempenho do professor, o que estaria refletida na unidimensionalidade encontrada. A correlação nota-pontuação foi estatisticamente significativa ($p<0,05$) em 12,2% dos casos, indicando interferência das notas de vários alunos na avaliação que estes fizeram de seus professores, mas, na maioria dos casos, não houve relação entre essas variáveis.

Palavras-chave: avaliação institucional; avaliação docente; validade e precisão; análise fatorial; análise de clusters.

Validity of a College Teacher Course Ratings Questionnaire

Abstract: The study investigates validity and reliability of a student's ratings questionnaire for the assessment of teacher effectiveness. It was composed by 15 questions with 5 point scale. A University at São Paulo State used it by asking students to respond on-line in Internet "secure mode". Participation was optional and voluntary. A total of 3689 students evaluated 568 teachers of 36 programs. The questionnaire was submitted to a factor analysis, cluster analysis and its scores were correlated student grades. Internal consistency reliability was excellent ($\alpha=0,965$). Items grouped in one factor. Cluster analysis confirmed this unidimensionality. Although this construct may be conceived as multidimensional, this finding suggests that students have a global opinion about teachers performance. The correlation grades-ratings was statistically significant ($p<0,05$) for 12% of cases. This indicates that although there were some cases where it was possible to infer some interference based on association of grades with teacher ratings in the majority of cases there is no such interference.

Key Words: teacher ratings; validity and reliability; factor analysis; cluster analysis.

* As atividades de pesquisa de Ricardo Primi que deram origem a este artigo são financiadas pelo CNPq e pela FAPESP. Correspondências devem ser enviadas para: Ricardo Primi, Universidade São Francisco, Laboratório de Avaliação Psicológica e Educacional (LabAPE), Mestrado em Psicologia, Rua Alexandre Rodrigues Barbosa, 45, CEP 13251-900, Itatiba, São Paulo, Fone (0XX11) 45348118, correio eletrônico: ricardo.primi@saofrancisco.com.br ou rprimi@uol.com.br. Os autores agradecem a Vania Franciscon Vieira, presidente da Comissão Própria de Avaliação da Universidade São Francisco pelo apoio e à Universidade São Francisco pelo fornecimento dos dados para esta pesquisa.

Introdução

Este estudo volta-se para a avaliação institucional dentro do sistema de ensino superior no Brasil, focalizando mais especificamente a avaliação do corpo docente por parte dos estudantes. Há quem argumente, dentro do meio acadêmico e das próprias instituições, que a avaliação feita pelos alunos seja um parâmetro inadequado para tomadas de decisão no âmbito acadêmico e administrativo. Isso porque os alunos seriam influenciados por fatores tais como: descompromisso com essa avaliação, interesses pessoais divergentes, qualidade do vínculo com o professor avaliado, qualidade das relações com as instâncias burocráticas, administrativas e financeiras da instituição, preocupação com as próprias notas ou conceitos nas disciplinas cursadas, entre outros.

Grenwald (1997) oferece uma revisão histórica sobre os estudos e preocupações relativas à validade e utilidade da avaliação dos estudantes em relação a seus professores. Na década de 1970, as pesquisas atribuíam à interferência das notas obtidas pelos estudantes um efeito negativo ou positivo sobre a avaliação dos professores, isto é, alunos que recebessem notas baixas poderiam, também, avaliar negativamente o professor, enquanto alunos com notas altas fariam o oposto. Se isso ocorrer como podemos nos certificar de que a avaliação está baseada na eficácia do desempenho do professor? Será que refletiria mais a experiência negativa que o estudante teve com esse professor? Essas são questões importantes concernentes à validade dessas avaliações. Na década de 1980, os estudos correlacionais, de modo geral, davam suporte à avaliação de docentes por estudantes, considerando que as eventuais interferências poderiam ser atribuídas a outros fatores, não necessariamente relacionados às notas escolares recebidas pelos estudantes (ex.: efetividade do ensino; motivação dos estudantes; interesse prévio do estudante pelo assunto, entre outros).

Na década de 1990, quatro tipos de questões compunham as principais discussões sobre o assunto: se as avaliações dos alunos seriam influenciadas por variáveis não relacionadas ao ensino efetivo; quão bem as avaliações se correlacionavam com outros indicadores de ensino efetivo; a unidimensionalidade ou multidimensionalidade da medida; e se as avaliações estariam sendo utilizadas para seus principais propósitos: avaliação de docentes e melhoria do ensino (Greenwald, 1997).

Essa última questão preocupava vários pesquisadores, e possivelmente continua sendo a principal preocupação nos dias de hoje. Armstrong (1998) teme que a avaliação de docentes por parte dos alunos, ao invés de favorecer a melhoria do ensino, termine por provocar o efeito contrário. Suas restrições dirigem-se à possibilidade de que essa avaliação produza um efeito colateral, levando professores a desinteressar-se em ajudar as pessoas a aprenderem, enquanto os aprendizes sentem-se menos responsáveis por sua própria aprendizagem, o que, no final das contas, seria danoso para o ambiente universitário. Buck (1998) compartilha dessas reflexões,

afirmando que o grande perigo é tornar os cursos divertidos e envolventes, embora vazios de conteúdo efetivamente capaz de preparar os alunos para o trabalho, que nem sempre é divertido, mas sempre exige compromisso.

Greenwald e Gillmore (1998) afirmam que o problema está na dúvida sobre se a avaliação feita pelo aluno baseia-se em comportamentos do professor que não contribuem para o aprendizado. Se isto for verdadeiro, esse tipo de avaliação não seria útil para a melhoria do ensino. Porém, sua prática continua em uso por falta de melhor alternativa. Em 1999, os mesmos autores consideram que as discussões a respeito da avaliação de docentes focam, em um extremo, positivamente a favor de sua validade e utilidade, e, no outro extremo, sua falta de validade e seu potencial danoso. Explícitamente, eles se posicionam no meio-termo, defendendo a premissa de que os problemas apresentados por esse tipo de avaliação podem ser corrigidos por meio estatístico, removendo os efeitos das interferências identificáveis. Uma das formas de proceder essa correção é sugerida introduzindo-se um sistema em que os escores médios são ajustados por meio de análise de regressão múltipla, que permite determinar a porcentagem da variância que não deveria ser interpretada como medida da qualidade instrucional (Gillmore e Greenwald, 1999).

Muitas pesquisas têm sido conduzidas no exterior a respeito da avaliação dos docentes pelos estudantes, tendo em vista preocupações quanto aos seus possíveis vieses.

Apesar disso, esses autores reconhecem a possibilidade do estabelecimento de um círculo vicioso não premeditado ou deliberado, mas como efeito colateral de um sistema de avaliação mal planejado e mal utilizado, no qual: as queixas dos alunos causam uma revisão nos conteúdos (para menos), a qual leva a um aumento nas notas acadêmicas (pela remoção de tópicos mais difíceis do conteúdo), que por sua vez levam a um aumento na pontuação na avaliação do docente (Gillmore e Greenwald, 1999).

Para Abrami e d'Apollonia (1999), seria pior desconsiderar a efetividade do ensino como critério para promoção e administração, pois se decisões precisam ser tomadas sobre a qualidade do ensino, necessita-se do apoio da melhor evidência existente, mesmo que ela não seja perfeita.

Losak (1994) adverte que os administradores de faculdades e universidades devem evitar focar somente testes e números; esconder os problemas e utilizar o auto-estudo como um instrumento promocional da instituição. Sobre este item, Howell e Symboluk (2001) investigaram as opiniões de alunos e professores de uma faculdade canadense sobre a divulgação das avaliações feitas por alunos sobre o ensino

oferecido pela escola. De modo geral, os alunos mostraram-se favoráveis a essa divulgação, listando uma série de vantagens, entre elas o acesso a informações sobre a instituição, que facilitaria a escolha de cursos e de professores. Os professores, apesar de concordarem com esses dois argumentos, mostraram-se desfavoráveis à divulgação dessas informações, preocupados com sua privacidade e com o efeito que uma avaliação negativa pode ter, mesmo que o professor já tenha modificado suas estratégias de ensino. Os autores sugerem como alternativa para essa divergência a criação de páginas institucionais na Internet, em que informações sobre cursos e professores sejam divulgadas sem comprometer a privacidade dos envolvidos. Além disso, sugerem tornar mais acessíveis as informações sobre os cursos antes da escolha pelos alunos ou publicar as diretrizes dos cursos antes do início da seleção, no começo de cada semestre ou ano.

Como se vê na literatura, a avaliação institucional e de docentes, com vistas a qualidade do ensino, é realizada em diversos países já há algum tempo. Hamdam Gonzáles (1996) afirma que a avaliação institucional das universidades é uma necessidade para se conseguir realizar mudanças exigidas pelo mundo contemporâneo, especialmente em países que pretendem sair de uma etapa de subdesenvolvimento.

Cashin (1996) argumenta que um sistema efetivo de avaliação institucional deve seguir certos princípios, tais como: clareza das metas institucionais; do propósito da informação a ser coletada para o desenvolvimento institucional, definição de suas fontes e de sua aceitabilidade técnica; envolvimento dos participantes no desenvolvimento do sistema pela comunicação aberta; apoio dos níveis administrativos superiores; flexibilidade e legalidade; responsabilidades do corpo docente e de outros membros da instituição; treinamento dos avaliadores e dos supervisores para darem o *feedback*; manutenção da confidencialidade; recompensa do desempenho efetivo; e revisão do sistema periodicamente.

A construção de um instrumento de auto-avaliação institucional, ou de avaliação de docentes – foco deste estudo – deve obedecer a requisitos técnicos exigidos de um instrumento de avaliação formal, para que decisões baseadas em seus resultados possam ser consideradas válidas e confiáveis. Esses requisitos referem-se a evidências de validade e precisão do instrumento, além de sua padronização.

Corcoran e Fisher (1987) chamam atenção para o fato de que validade e precisão de um instrumento devem ser revistas periodicamente, para que os dados sejam atualizados e se tornem aplicáveis à população atual. Eles também recomendam a escolha de instrumentos cujo coeficiente de precisão seja da ordem de 0,80, pelo menos, e cuja validade tenha sido estabelecida por diversos estudos e não apenas por um único. Além destes critérios, os autores recomendam a observação de mais três aspectos ao selecionar um instrumento de avaliação para aplicação: a extensão do instrumento, a aceitabilidade social de seu conteúdo e a habilidade do aplicando para responder a ele (facilidade de compreensão dos itens propostos).

Em conformidade com as considerações de Corcoran e Fischer (1987), um instrumento de avaliação de desempenho docente para ser usado em um sistema amplo de auto-avaliação institucional deve receber atenção quanto aos requisitos técnicos, sobretudo quando decisões importantes podem ser tomadas com base nos resultados obtidos por esse instrumento.

Auto-avaliação institucional e avaliação de docentes

Um dos aspectos avaliados por escolas, faculdades, universidades e sistemas educacionais, ao realizar uma avaliação institucional, em diversos países, é o desempenho dos docentes, considerado um indicador da qualidade do ensino.

Nos EUA, a Secretaria da Educação do Estado de Utah (Utah State Office of Education, 1994) estabeleceu um sistema de avaliação docente baseado em aprovação de programa. Esse sistema avaliativo difere do enfoque voltado para a avaliação dos professores por parte dos alunos, estratégia bastante utilizada e estudada nos últimos anos.

O Pima Community College (1994), no Arizona (EUA), utiliza um processo de avaliação institucional em intervalos de 5 anos, porém o corpo docente é avaliado a cada 3 anos por duas classes selecionadas da carga regular de aulas do membro docente, utilizando um formulário de avaliação cumulativa dos docentes.

Muitas pesquisas têm sido conduzidas no exterior a respeito da avaliação dos docentes pelos estudantes, tendo em vista preocupações quanto aos seus possíveis vieses. Basow e Silberg (1987) analisaram a avaliação de docentes realizada por estudantes para investigar a influência do gênero sobre os resultados. Dezesesseis professores de ambos os sexos foram avaliados por cerca de 1000 estudantes de ambos os sexos, de uma faculdade particular no interior dos EUA, por meio de dois instrumentos: o *Teacher Rating Form* (26 itens em escala Likert de 5 pontos, variando de bem acima da média (1) a bem abaixo da média (5), sendo que a primeira questão tratava da habilidade geral de ensino e a escala de 5 pontos variava de notável (1) a pobre e inadequada (5). As demais questões eram igualmente distribuídas entre 5 fatores: escolaridade, organização/clareza, interação instrutor-grupo, interação instrutor-aluno individualmente, dinamismo/entusiasmo, e o segundo instrumento, a forma reduzida do *Bem Sex-Role Inventory* – BSRI -, que foi utilizado para avaliar como os alunos percebiam os traços masculinos ou femininos de seus professores. Apesar da pequena amostra e da região aparentemente mais tradicional onde o estudo foi realizado, os resultados foram significativos, o que sugere cuidado ao levar em conta as avaliações de docentes por parte de estudantes pois pode ocorrer efeito de interação em função do gênero.

Erdle, Murray e Rushton (1986) estudaram a validade da avaliação dos docentes por estudantes por meio da relação entre as dimensões de personalidade e de

comportamento instrucional, usadas como preditores, e as avaliações feitas pelos alunos em relação aos professores usadas como critério. Por meio de *path analysis* (os coeficientes são pesos beta padronizados derivados de uma série de análises de regressão múltipla), chegaram à conclusão de que a personalidade influencia a efetividade instrucional via comportamento em sala de aula. Assim, aparentemente, professores com alto nível de Orientação Interpessoal envolvem-se em frequências altas de comportamentos carismáticos e organizados em classe, os quais, por sua vez, levam a avaliações positivas dos alunos.

Erdle, Murray e Rushton (1986) concluíram que a personalidade do professor reflete-se em certos comportamentos em sala de aula, os quais são avaliados de forma válida pelos estudantes. Essa evidência contraria a posição, defendida por muitos, de que as avaliações realizadas pelos alunos não têm validade, porque seriam influenciadas fortemente pela personalidade do professor. Para os autores do estudo, o comportamento em classe é o mediador da relação entre personalidade e efetividade de ensino do professor, sendo que os alunos avaliam seus professores em relação aos comportamentos que eles apresentam na classe durante as aulas.

Como pode ser notado nessa breve revisão da literatura ainda existem muitas questões em aberto sobre a validade dos instrumentos que pretendem avaliar o desempenho do professor com base na opinião do aluno. Apesar disso, várias universidades no exterior e no Brasil já empregam instrumentos desse tipo em sua sistemática de avaliação institucional. Diante desse cenário, este estudo pretende oferecer uma contribuição apresentando uma pesquisa de validade de um instrumento de avaliação docente. Espera-se que as informações levantadas aqui possam contribuir para uma interpretação mais adequada desses instrumentos.

Problema, objetivos e hipóteses

A questão principal que os instrumentos de avaliação docente suscitam é: “O que esses instrumentos medem de fato?” Dessa questão geral surgem algumas perguntas mais específicas, tais como, se existem várias dimensões que devem ser consideradas para abarcar todos os aspectos do desempenho de docente, já que os docentes podem ter um bom desempenho em um aspecto, mas não em outro; e se os alunos conseguem fazer uma avaliação do docente isenta de influências da avaliação que o professor faz deles, especialmente quando apresentam baixo desempenho (notas).

Partindo dessas questões, este estudo foi esboçado com o propósito de buscar evidências de validade de um instrumento de avaliação docente empregado em uma universidade do interior paulista, especialmente com dois objetivos: (a) investigar a validade de construto utilizando análise fatorial dos itens e análise de perfis. Esta última, para verificar se a estrutura correlacional entre os itens é condizente com um modelo multidimensional que foi subjacentemente pressuposto na sua construção; e

(b) a validade de construto convergente-discriminante, investigando as associações entre as notas que o aluno atribui ao professor com as notas que recebe deste nas avaliações de aprendizagem. Pretende-se com este último objetivo investigar em que medida essas duas variáveis são dependentes, para se discutir aspectos relacionados à variância confiável de construto irrelevante, uma vez que tal associação teoricamente não deveria existir, pois a avaliação do docente e o desempenho escolar, embora relacionados, são construtos diferentes e suas medidas deveriam divergir.

Método

Participantes:

Este estudo considerou os dados das avaliações efetuadas por 3689 estudantes, avaliando 568 professores de 36 cursos de graduação e pós-graduação, distribuídos em quatro campi da universidade. Cada aluno avaliava mais de um professor na disciplina específica que eles ministravam. Os professores se distribuíam em 720 disciplinas, sendo que algumas delas eram ministradas por mais de um professor (trabalho de conclusão de curso, por exemplo). Em média, cada aluno respondia sete questionários, um para cada professor. Combinando o número de alunos com o número de questionários respondidos, obtêm-se um total de 25387 avaliações. No sistema adotado pela instituição, a resposta à avaliação é espontânea. Em geral a taxa de participação gira em torno de 25%. Se considerarmos o número de questionários completos respondidos para cada professor, observamos um mínimo de um questionário e máximo de 228 questionários, com média de 45 avaliações por professor.

Material

Foi utilizado um instrumento de avaliação institucional, com itens relativos à qualidade dos serviços prestados pela instituição, dos quais 15 itens referiam-se à qualidade do desempenho dos docentes, em escala Likert variando de 1 (péssimo) a 5 pontos (excelente). Na Tabela 1, na coluna nomeada “Item” pode-se ver os quinze itens.

Procedimento

O instrumento foi disponibilizado a todos os alunos da graduação e pós-graduação, no período de 30 de outubro a 13 de dezembro de 2004, em meio eletrônico, via Internet, em página que só pode ser acessada com identificação numérica e senha individual do aluno, independentemente do local em que ele esteja realizando o acesso. Os alunos receberam e-mail personalizado convidando-os a realizarem a avaliação. Foi colocada uma mensagem na página para alunos, além de notícia na página inicial do portal, cartazes em salas de aulas e corredores, e aviso oral pelos representantes de classe.

Apesar da ampla divulgação, a participação do aluno é voluntária. Ao decidir participar, o aluno acessa a ligação existente na página do aluno e passa a responder os itens relativos ao professor que aparece identificado e que ministrou aulas naquele ano para aquele aluno. Os itens vão aparecendo à medida que são respondidos. Quando termina a avaliação desse professor, aparece na tela do computador o nome de outro professor para sua avaliação, e assim sucessivamente, até que todos os professores da série em que o aluno está matriculado tenham sido avaliados.

Resultados e Discussão

Na Tabela 1 são apresentadas as estatísticas descritivas dos itens do instrumento avaliando os aspectos do desempenho docente. A pontuação de cada item varia de 1 (péssimo), 2 (fraco), 3 (regular), 4 (bom) e 5 (excelente). Os resultados obtidos evidenciaram que a menor média de pontuação ocorreu para o Item 3 e a maior média para o Item 4 da escala, o que parece indicar que, a despeito do conhecimento evidenciado pelo professor, a dificuldade reside na implementação da tão propalada interdisciplinaridade. De maneira geral a média pode ser considerada alta já que em todos os itens estão muito próximas de 4 (bom).

Além disso, chama atenção o número de respostas não computadas automaticamente (perdas) para o Item 6, que chega a 1/3 do número de participantes. Esse fato pode estar relacionado a características específicas de cada curso, já que nesta análise estão sendo considerados os dados globais obtidos pela avaliação da Universidade como um todo. Uma possibilidade a ser considerada é que talvez nem todos os cursos possuam atividades do tipo: supervisão de estágio ou de prática, ou mesmo TCC ou Relatório Final, daí o elevado número de omissões ou respostas nulas.

A análise global dos dados obtidos foi realizada em duas fases, a saber: a primeira fase teve como objetivo a análise do instrumento empregando-se a análise fatorial exploratória, investigando sua validade de construto, para responder a primeira questão levantada sobre o número de dimensões necessárias para a avaliação do docente. Também se empregou um método alternativo de análise de perfis para se investigar a utilidade diagnóstica de um modelo multi ou unidimensional. A segunda fase teve por objetivo investigar as relações entre as avaliações dos docentes e o desempenho dos estudantes nas disciplinas, para investigar a segunda questão: Em que medida as avaliações estão correlacionadas com as avaliações do desempenho acadêmico?.

Análise fatorial: a estrutura interna dos itens

A análise fatorial por componentes principais da matriz de correlações entre os itens da escala de avaliação docente, apresentada na Tabela 2, foi realizada utilizando-se o SPSS (*Statistical Package for Social Sciences*). A decisão sobre o número de

Tabela 1. *Estatísticas descritivas e resultados da análise fatorial*

Item	M	DP	N Válidos	Cargas Fatoriais	Comunalid
1. a clareza das explicações do professor nos assuntos tratados?	4,04	1,089	25267	0,862	0,743
2. a relação que o professor estabelece entre os aspectos conceituais da disciplina e as aplicações práticas?	4,07	1,052	25167	0,862	0,743
3. a integração que o professor faz desta disciplina com as outras disciplinas do curso (interdisciplinaridade)?	3,92	1,062	24720	0,819	0,671
4. o conhecimento que o professor tem dos conteúdos desta disciplina?	4,51	0,842	25285	0,752	0,565
5. a organização das aulas práticas?	4,00	1,111	22858	0,845	0,715
6. a atenção dispensada à orientação/supervisão de Estágios, Prática Jurídica, TCC ou Relatório Final?	4,02	1,078	16655	0,805	0,648
7. o aproveitamento do tempo destinado às aulas?	4,10	1,063	24541	0,816	0,665
8. o cumprimento do programa apresentado?	4,22	0,964	23997	0,809	0,654
9. a adequação dos métodos de ensino utilizados pelo professor?	4,00	1,106	23982	0,875	0,766
10. a adequação das avaliações em relação aos conteúdos trabalhados e discutidos na disciplina?	4,10	1,044	23039	0,810	0,656
11. a clareza/objetividade dos critérios de avaliação usados pelo professor?	4,07	1,056	22806	0,811	0,658
12. a análise que o professor faz dos resultados das avaliações visando melhoria da aprendizagem?	3,99	1,061	21741	0,816	0,666
13. a motivação/incentivo dado pelo professor para você participar das aulas?	3,94	1,201	22926	0,831	0,690
14. a atenção do professor para responder as dúvidas?	4,31	1,005	22989	0,789	0,622
15. a postura ética do professor (respeito aos alunos, professores, à profissão e à Instituição)?	4,49	0,885	22910	0,680	0,463

fatores a serem extraídos foi tomada analisando-se o gráfico *Scree*, o qual indicou a presença de um único fator que explica 66,167 da variância total da escala, mais outros fatores residuais. O auto-valor do primeiro fator foi 9,92, enquanto o do segundo foi 0,72, indicando claramente que um único fator explica as correlações entre os quinze itens já que o critério alternativo de reter fatores com autovalores maiores que 1 também indica a existência de um único fator.

Tabela 2. Matriz de correlação entre os itens do instrumento de avaliação docente.

ITENS	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2.	0,828													
3.	0,718	0,741												
4.	0,668	0,658	0,599											
5.	0,727	0,730	0,693	0,610										
6.	0,642	0,654	0,661	0,573	0,669									
7.	0,681	0,670	0,636	0,604	0,702	0,614								
8.	0,655	0,658	0,624	0,606	0,679	0,616	0,725							
9.	0,765	0,753	0,701	0,617	0,754	0,665	0,708	0,695						
10.	0,647	0,650	0,609	0,549	0,643	0,614	0,618	0,635	0,695					
11.	0,649	0,648	0,609	0,540	0,637	0,629	0,607	0,621	0,686	0,759				
12.	0,638	0,642	0,633	0,536	0,640	0,642	0,621	0,624	0,689	0,697	0,714			
13.	0,704	0,692	0,664	0,550	0,682	0,644	0,647	0,611	0,732	0,633	0,642	0,686		
14.	0,632	0,630	0,581	0,583	0,605	0,640	0,586	0,588	0,640	0,604	0,612	0,635	0,676	
15.	0,517	0,511	0,491	0,504	0,496	0,554	0,510	0,518	0,527	0,514	0,524	0,547	0,545	0,636

As últimas duas colunas da Tabela 1 apresentam as cargas fatoriais dos itens no único fator extraído e a comunalidade, respectivamente. Como pode ser observado, as cargas (correlação item – fator) no fator extraído, são altas, variando de 0,87 (item 9) a 0,68 (item 15). De forma correspondente, as comunalidades foram altas, pois os valores variaram de 0,76 (item 9) a 0,46 (item 15), indicando que a variância no item tende a ser explicada por esse único fator. A análise da consistência interna da escala pelo coeficiente alfa de Cronbach aponta valor elevado de $\alpha = 0,965$, conforme recomendado pela literatura, o que indica um nível excelente de consistência interna do instrumento.

Esses resultados indicam que um modelo unidimensional do desempenho docente, quando avaliado pelo aluno, explica as relações entre os itens. Embora a literatura discuta o desempenho docente como um construto multidimensional (Greenwald, 1997) composto por várias áreas e os itens tenham sido construídos com base nessa concepção, os resultados da análise fatorial indicam a existência de um único fator, isto é, uma solução unidimensional. Uma possível interpretação para este fato é que, embora esse construto possa ser multidimensional, na concepção do aluno ele é unidimensional. Parece que o aluno tem uma opinião geral do desempenho do docente e atribui essa nota aos vários indicadores sem fazer uma distinção mais fina entre eles. Esse resultado é condizente com o que tem sido relatado na literatura como, por exemplo, Cashin e Downey (1992). Isso sugere que o instrumento pode ser reduzido, tornando-se mais econômico sem perder as informações que permite coletar. Como resposta à pergunta “o que esse instrumento mede”, podemos ao menos dizer que ele mede uma opinião geral que o aluno forma do desempenho do docente.

Análise de clusters: a informação diagnóstica por intermédio da análise de perfis

Quando instrumentos desse tipo são usados, espera-se que seus resultados possam conseguir indicar eventuais áreas de menor desempenho para que se possa pensar em medidas de treinamento para aprimorar as fraquezas. Essa idéia também é condizente com uma concepção multidimensional do desempenho já que pressupõe uma independência de diferentes aspectos de modo que o professor poderia ter desempenho fraco em uma área, mas bom em outras. A análise fatorial indica uma solução unidimensional e, sendo assim, as diferentes áreas estão correlacionadas. Como decorrência, na opinião do aluno, se o professor apresenta um desempenho bom em uma área ele tende a apresentar o mesmo nível nas outras áreas. Nesse sentido a informação diagnóstica fica limitada àquela presente no escore geral, que indica a avaliação global que o aluno faz do desempenho geral do professor. Não haveria sentido considerar escores em áreas específicas já que eles tendem a estar associados trazendo a mesma informação.

Uma forma de investigar os padrões revelados pelo instrumento indicando uma riqueza diagnóstica é a análise de *clusters* (Hair, Anderson, Tatham & Black, 2005). Essa análise investiga os perfis obtidos pelos professores e os agrupa em razão da semelhança, constituindo grupos de professores com avaliações semelhantes em todo instrumento. A análise desses grupos pode indicar se existem grupos com perfis diferenciados decorrentes de uma configuração de itens com avaliações altas contrastando com itens com avaliações baixas, padrões esses, mais associados a uma interpretação multidimensional; ou se existem grupos mais condizentes com a concepção unidimensional, em que só diferem em função do nível geral das avaliações. Esse método de análise foi adotado como uma forma de escrutinar novas possibilidades de multidimensionalidade nos dados.

Para essa análise, os dados foram reorganizados de acordo com professores e disciplinas. Foram selecionados os professores que obtiveram pelo menos quinze avaliações por disciplina que ministravam. Para cada combinação professor-disciplina, foram calculadas as médias em cada item do questionário. Foram analisados 617 casos, os quais incluem um número menor de professores, porque um professor poderia ter ministrado mais de uma disciplina. Empregou-se a análise de *cluster* para agrupar casos com perfis semelhantes de média de desempenho nos quinze itens avaliados. O procedimento de análise calcula uma matriz de distância euclidiana ao quadrado entre cada par possível de casos e utiliza um algoritmo de aglomeração para, passo a passo, ir formando grupos de casos com perfis mais semelhantes possíveis. Utilizou-se a fórmula de Ward, baseada na análise de variância, como algoritmo de aglomeração. Uma análise visual do dendograma sugeriu a composição de dez grupos.

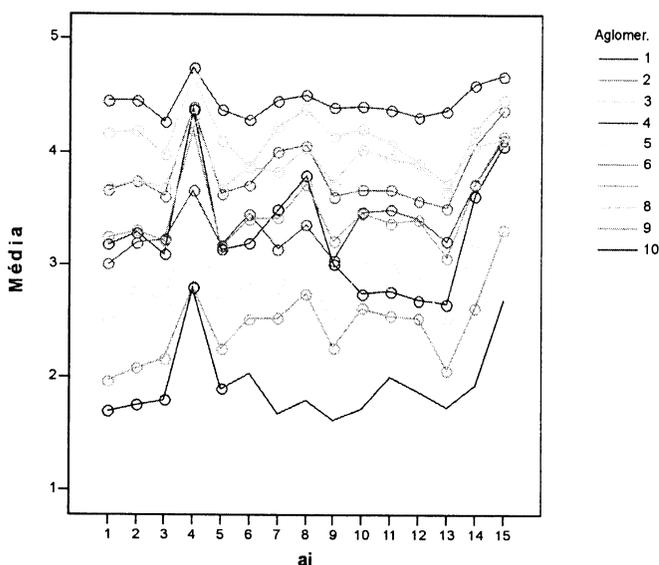
Para verificar se esses aglomerados captam diferenças sistemáticas entre as médias foi feita uma ANOVA 15 X 10, considerando 15 medidas repetidas correspondentes aos quinze itens, e dez subgrupos de sujeitos, sendo essas as duas variáveis independentes. A variável dependente era a avaliação na escala de 1 (péssimo) a 5 (excelente). Os resultados são apresentados na Tabela 3 e os perfis de médias na Figura 1. Essa figura mostra no eixo vertical a escala de pontos usada na avaliação do professor (a variável dependente) e no eixo horizontal quinze pontos correspondentes aos itens do instrumento. Os dez perfis coloridos (aglomerados) correspondem às médias dos sujeitos alocados em cada um dos dez subgrupos nos quinze itens do instrumento.

Como pode ser observado na Tabela 3, o efeito dos itens foi significativo ($F=201,6$) indicando existir diferenças significativas globais entre as médias nos itens do instrumento, independentemente do subgrupo de sujeitos formados a partir da análise de *cluster*. Isso decorre do fato de que em alguns itens os professores, em geral, são mais bem avaliados do que em outros, como já se comentou. Também se observou um efeito importante de diferenças gerais entre os dez subgrupos indepen-

Tabela 3. *Resultados da ANOVA 15 X 10 (15 medidas repetidas dos quinze indicadores e dez subgrupos de sujeitos).*

Efeito intra-Itens					
Fonte	TIPO III Soma dos quadrados	gl	MÉDIA	F	Sig.
Itens	107,748	14	7,696	201,67	0,000
Itens X aglomerados	112,843	126	0,896	23,46	0,000
Erro (Itens)	324,299	8498	0,038		
Efeito entre Grupo					
Aglomerados	2026,444	9	225,16	538,9	0,000
Erro	253,603	607	0,418		

dentemente dos itens, isto é, uma diferença acentuada no escore global calculado pela soma dos quinze itens, entre os subgrupos de sujeitos. Aliás, esse foi o maior efeito observado nessa análise, como indicado no efeito “Aglomerado” ($F= 538,9$). Por último, observou-se uma interação significativa entre o perfil de médias nos itens com os subgrupos (Itens X Aglomerado), indicando que o desenho das linhas no gráfico (Figura 1), com altos e baixos nos quinze itens, não é exatamente semelhante para todos os dez subgrupos. Embora esse efeito tenha sido significativo, foi o de menor magnitude em comparação com os outros dois ($F=23,4$).

Figura 1. *Perfis dos subgrupos de professores com avaliações diferenciadas obtidos na análise de clusters.*

Esses efeitos ficam mais claros na Figura 1. Uma parte mais notável desses padrões refere-se à organização dos perfis dos grupos por nível, isto é, os subgrupos se diferenciam, principalmente, pela posição na escala vertical. Por exemplo, os Subgrupos 9 e 10 nos níveis 1 e 2, os Subgrupos 1 e 3 nos níveis 4 e 5. Esse padrão é mais condizente com a hipótese de unidimensionalidade. Em um padrão multidimensional, caracterizado por itens com uma relativa independência, os subgrupos deveriam apresentar configurações caracterizadas por altos e baixos em itens diferentes, padrão este que sugeriria uma organização independente dos itens. Portanto, nessa análise da informação diagnóstica fica mais uma vez evidente que o escore global do instrumento apreende grande parte da informação de desempenho observada por esse instrumento. Portanto, grande parte da variância está associada ao nível global (aglomerados) indicando unidimensionalidade, e não aos desenhos de perfis os quais indicariam multidimensionalidade (aglomerados X itens).

Análise correlacional entre desempenho do docente e notas escolares

Um último aspecto investigado foi o das associações entre as notas que o aluno atribui ao professor com as notas que recebe deste nas avaliações de aprendizagem. Investigou-se em que magnitude essas duas medidas estão associadas para se discutir questões de validade convergente-divergente. Teoricamente, não deveria existir associação entre essas medidas porque os construtos avaliação do desempenho escolar e avaliação do docente, embora relacionados, são diferentes e, portanto, suas medidas deveriam divergir. A análise da dependência entre essas duas variáveis deve contribuir para aprimorar a interpretação dos escores do teste à luz dos conceitos de variância confiável de construto irrelevante.

Para essa análise, foram calculadas, para cada combinação aluno/professor, duas variáveis. A primeira indicava a avaliação que o aluno fazia do professor, calculada pela média de pontos nos 15 itens avaliados. A segunda foi calculada pela média final do desempenho do aluno naquela disciplina ministrada pelo professor. As correlações entre essas duas variáveis foram calculadas para cada combinação professor/disciplina com o intuito de explorar se ocorria, em cada evento específico – caracterizado por um professor x , ministrando uma disciplina y para um conjunto de n alunos – uma correlação entre o desempenho dos alunos e a avaliação que faziam do professor.

Foi criado um banco de 1700 combinações professor/disciplina, mas não foi possível calcular as correlações para todas elas pela ausência de informações completas, por exemplo, nota do aluno e/ou falta de número mínimo de alunos que responderam à avaliação de um professor em uma disciplina específica. Foram considerados os casos em que existiam pelo menos dez avaliações, o que resultou em 768 casos. Desses casos, foram encontrados 84 (10,9%) coeficientes de correlação posi-

tivos significativos ($p < 0,05$); 10 (1,3%) casos de coeficientes de correlação negativos significativos e 674 (87,8%) coeficientes não significativos. Embora o número de coeficientes indicando associação não seja alto, eles não podem ser negligenciados, uma vez que, em número absoluto, superam o índice tradicional de 5% de coeficientes significativos ao acaso.

Na figura 2, apresenta-se a distribuição desses coeficientes e na Tabela 4 as estatísticas descritivas. A distribuição é aproximadamente normal centrada em 0,10, isto é, um pouco acima de zero.

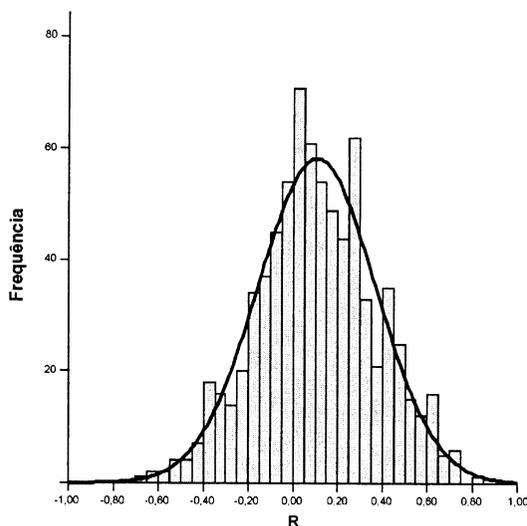


Figura 2. *Distribuição dos coeficientes de correlação nota X avaliação do professor.*

Tabela 4. *Estatísticas descritivas dos coeficientes de correlação nota X avaliação do professor.*

	Estatísticas	Erro Padrão
N válido	768	
Mínimo	-0,695	
Máximo	0,830	
Média	0,106	
Desvio Padrão	0,263	
Assimetria	-0,025	0,088
Curtose	-0,148	0,176

Para explorar melhor essas associações, foram feitas mais algumas análises comparativas dos subgrupos nos quais foram observadas associações significativas. Para isso, foram separados três grupos: o Grupo 1 (G1) formado por alunos com o padrão de correlações positivas significativas entre a avaliação do professor e nota, Grupo 2 (G2) formado por alunos com o padrão de correlações negativas significativas entre a avaliação do professor e nota e Grupo 3 (G3) no qual não houve associação significativa. Para cada combinação docente/disciplina calcularam-se as médias e desvios-padrão do desempenho do professor e das notas dos alunos na disciplina. Por fim, foram efetuadas quatro ANOVAs comparando os resultados médios dos três grupos nessas quatro variáveis. Os resultados estão apresentados na Tabela 5.

Tabela 5. Resultados de quatro ANOVAs comparando médias de desempenho e de avaliação do professor para grupos com ou sem correlação entre essas duas variáveis.

ANOVA	Desempenho do Professor (1-5)		Nota do Aluno (0-10)	
	M	DP	M	DP
F	1,62	***13,39	***8,44	***30,14
Sig	0,19	0,000	0,000	0,000
gl	2	2	2	2
Média ao Quadrado	0,42	1,07	9,40	12,92
Grupos	M	DP	M	DP
G1 $p < 0,05$ $r +$				
N	102	102	102	102
M	4,01	0,79	7,04	1,71
DP	0,48	0,29	1,04	0,76
G2 $p < 0,05$ $r -$				
N	21	21	21	21
M	4,10	0,57	7,95	1,07
DP	0,67	0,31	1,13	0,70
G3 $p > 0,05$				
N	1018	1018	1018	1018
M	4,10	0,64	7,40	1,19
DP	0,51	0,28	1,05	0,64
Total				
N	1141	1141	1141	1141
M	4,09	0,65	7,38	1,24
DP	0,51	0,28	1,06	0,67

Como pode ser observado, o desempenho médio dos professores do G1, G2 e G3 não difere significativamente. Já o desvio padrão desse desempenho difere. No Grupo 1, há uma maior dispersão das avaliações entre os alunos, isto é, as notas atribuídas aos professores variam mais entre os alunos do subgrupo em que se observa uma associação positiva com a nota. Um fato interessante refere-se à variação do desempenho dos alunos, uma vez que o desempenho acadêmico dos alunos do subgrupo em que houve correlação avaliação-nota (G1) difere significativamente do subgrupo em que não houve essa correlação (G3). No primeiro grupo o desempenho médio tende a ser menor e mais variável entre os alunos.

Conclusão

A análise fatorial e a análise de clusters indicaram que o Instrumento de Avaliação de Docentes do Ensino Superior utilizado nesta Universidade do interior do Estado de São Paulo, ao contrário das expectativas, apresenta unidimensionalidade. Esse dado não resolve a discussão existente na literatura quanto à multidimensionalidade ou unidimensionalidade da medida do construto, mas fornece indicação de que, para o aluno, a despeito dos diversos fatores que envolvem o desempenho do professor, sua compreensão dos itens indica uma opinião geral do desempenho.

Assim, no que se refere à validade do instrumento, isto é, se ele efetivamente avalia aquilo a que se propõe, pode-se dizer que ele avalia uma opinião geral que o aluno forma a respeito do desempenho de seus professores. Sendo assim, o instrumento pode ser reduzido, sem perder suas qualidades, tornando-se mais econômico tanto em extensão como em tempo necessário para ser respondido uma vez que não adianta torna-lo demasiado longo a pretexto de abarcar mais dimensões do desempenho se de fato os alunos não respondem considerando tais nuances. No que se refere à precisão desse score, o alfa de Cronbach evidenciou excelente consistência interna atingindo um nível de $\alpha = 0,965$.

Com relação à dúvida sobre a influência das notas obtidas pelos alunos na avaliação que eles fazem de seus professores, o estudo correlacional evidenciou que em 12,2% dos casos houve associação entre essas variáveis. Embora reduzido, esse número não pode ser desprezado porque se mostrou estatisticamente significativo a $p < 0,05$. Observou-se maior variação na pontuação atribuída aos professores no grupo de alunos em que houve associação positiva entre as duas variáveis. Curiosamente, esse grupo difere significativamente em termos de desempenho acadêmico médio do grupo em que não houve correlação entre as variáveis, isto é, sua média acadêmica é menor e mais variável entre os alunos. Isso significa que pode haver interferência das notas recebidas pelo aluno na avaliação que este faz de seus professores, mas, na maioria dos casos, não há relação entre essas variáveis. Portanto essa variável deve ser considerada caso a caso para que, caso sejam encontradas associações entre a avaliação do professor com o desempenho sejam efetuadas as devidas ponderações.

Referências

ABRAMI, P. C. e D'APOLLONIA, S. Current concerns are past concerns. **American Psychologist**, v. 54, p. 519-520, jul. 1999

ARMSTRONG, J. S. Are student ratings of instruction useful? **American Psychologist**, v. 53, n. 11, p. 1223-1224. 1998

ASSESSORIA DE COMUNICAÇÃO SOCIAL. Ministério da Educação e Cultura. Sistema Nacional de Avaliação e Progresso da Educação Superior. Uma nova sistemática de avaliação do Ensino Superior Brasileiro. **Disponível no endereço: <<http://www.mec.gov.br>>. Acesso: 28 jan 2004.**

BASOW, S. A. e SILBERG, N. T. Student evaluations of college professors: Are female and male professors rated differently? **Journal of Educational Psychology**, v. 79, n. 3, p. 308-314. 1987

BRASIL. Decreto Nº 3.860, de 9 de julho de 2001. Dispõe sobre a organização do ensino superior, a avaliação de cursos e instituições, e dá outras providências. **Diário Oficial da União.** 10 jul 2001.

BUCK, D. Student evaluations of teaching measure the intervention, not the effect. **American Psychologist**, v. 53, n. 11, p. 1224-1225. 1998

CASHIN, W. E. Developing an effective faculty evaluation system. **IDEA Paper.** Kansas State University, Center for Faculty Evaluation and Development, Division of Continuing Education. 1996. Nº. 33. Abstract from ERIC database. ERIC identifier 395536. Disponível em <<http://SearchERIC.org/ericdb/ED395536.htm>> Acesso 15 jan 2004.

CASHIN, W. E., e DOWNEY, R. G. Using global student ratings for summative evaluation. **Journal of Educational Psychology**, v. 84, p. 563-572. 1992

CORCORAN, K. e FISHER, J. **Measure for clinical practice: a sourcebook.** New York, USA: Free Press, 1987

CRONBACH, L. J. **Fundamentos da testagem psicológica.** 5ª. ed., Porto Alegre: Artes Médicas, 1996.

ERDLE, S., MURRAY, H. G., e RUSHTON, J. P. Personality, classroom behavior, and student ratings of college teaching effectiveness: a path analysis. **Journal of Educational Psychology**, v. 77, n. 4, p. 394-407. 1985

GILLMORE, G. M. e GREENWALD, G. M. Using statistical adjustment to reduce biases in student ratings. **American Psychologist**, v. 54, p. 518-519, jul. 1999

GREENWALD, A. G. e GILLMORE, G. M. How useful are student ratings? Reactions to comments on the Current Issues Section. **American Psychologist**, v. 53, n. 11, p. 1228-1229. 1998

GREENWALD, A. G. Validity concerns and usefulness of student ratings of instruction. **American Psychologist**, v. 52, n. 1, p. 1182-1186. 1997

HAIR, J. F.; ANDERSON, R. E. ; TATHAM, R. L.; e BLACK; W. C. **Análise multivariada de dados**. 5ª. ed., Porto Alegre: Artes Médicas, 2005.

HAMDAN GONZÁLEZ, N. Evaluación institucional de las universidades. **Revista de la Facultad de Medicina**, v. 19, n. 2, p.155-158. 1996.

HOWELL, A. J. e SYMBALUK, D. G. Published student ratings of instruction: revealing and reconciling the views of students and faculty. **Journal of Educational Psychology**, v. 93, n. 4, p. 790-796. 2001.

LOSAK, J. Issues/problems in assessing and documenting institutional effectiveness. research and planning report 94-01. **Southern Association of Colleges and Schools, 1994. Abstract from ERIC database. ERIC identifier ED403812. Disponível no endereço: <<http://SearchERIC.org/ericdb/ED403812.htm>> Acesso 15 jan. 2004.**

PIMA COMMUNITY COLLEGE, Office of the Vice Chancellor. Board and employee evaluation. **Pima County Community College. District Institutional Effectiveness. Arizona – USA: Pima Community College, 1994. 34 p. Series 4. Abstract from the ERIC database. ERIC Identifier: ED375915. Disponível no endereço: <<http://SearchERIC.org/ericdb/ED375915.htm>>. Acesso: 15 jan. 2004.**

UTAH STATE OFFICE OF EDUCATION. **Procedures for the Evaluation of Teacher Education Programs**. Feb. 1994.33p.