

LIMITE DA VALIDADE DE UM INSTRUMENTO DE AVALIAÇÃO DOCENTE

Cristiano Mauro Assis Gomes¹ - Universidade Federal de Minas Gerais
Oto Borges - Universidade Federal de Minas Gerais

RESUMO

Este trabalho descreve e discute as qualidades psicométricas do instrumento de avaliação docente de uma instituição de ensino superior. As médias dos escores atribuídos pelos estudantes aos professores da instituição foram analisadas pelas seguintes técnicas estatísticas: análise fatorial exploratória, método dos eixos principais e análise paralela por permutação. Os resultados indicam que um único fator explica 78,8% da variância comum dos itens, e que o coeficiente alfa de Cronbach é 0,98. Os resultados indicam um único fator explicando 78,8% da variância comum dos itens. O coeficiente alfa de Cronbach encontrado foi 0,98. Tais resultados indicaram um instrumento confiável e unidimensional que acessa a avaliação socialmente compartilhada pelos estudantes concernente ao seu grau de satisfação em relação ao trabalho docente. Como uma das implicações, destacamos a necessidade de elaboração de um instrumento de avaliação docente mais adequado, capaz de identificar uma ampla variedade de ações e comportamentos docentes dando ênfase na percepção delas pelos estudantes e não na avaliação dos docentes.

Palavras-chave: avaliação docente; avaliação educacional; validade.

VALIDITY LIMIT OF A FACULTY STAFF EVALUATION INSTRUMENT

ABSTRACT

The present work describes and discusses the psychometric qualities of a faculty staff evaluation instrument used on an higher education institution. The score averages attributed by students to their lecturers were subject to three statistical techniques: exploratory factorial analysis, principal axis factoring and parallel permutations analysis. Results indicate that a single factor explains 78.8% of common item variance and Cronbach's alpha coefficient is 0,98. Such results indicate that the faculty staff evaluation instrument is reliable. The instrument is one dimensional and capable of revealing socially shared students' levels of satisfaction regarding the work of faculty members. As an implication, we highlight the necessity of elaborating a more adequate instrument, capable of identifying a large variety of actions and teaching behaviors, with emphasis on the students' perceptions of such aspects instead of teachers' evaluation.

Keywords: teaching evaluation; educational evaluation; validity.

INTRODUÇÃO

Desde o final da década de 1980 as políticas educacionais brasileiras têm enfatizado a importância da avaliação institucional e incentivado a implementação de processos endógenos, conduzidos por equipes internas às instituições, e processos exógenos, que se baseiam em mecanismos de avaliação externos à instituição. De uma maneira geral, um dos componentes mais usuais e constantes dos processos avaliativos endógenos é a avaliação do trabalho docente pelo aluno. Os instrumentos de avaliação docente pelo discente são bastante parecidos. Eles são compostos por itens objetivos que procuram mensurar as diversas dimensões do trabalho docente, tais como aspectos didáticos, a competência na gestão da sala de aula, a capacidade de interação com a turma, entre outros aspectos. Cada item do instrumento é respondido através de uma escala do tipo *likert*. Em

cada item, o estudante escolhe qual das categorias disponibilizadas na resposta melhor expressa a sua avaliação a respeito da dimensão específica trabalho docente ao qual o item se refere.

Neste trabalho, questionamos a validade de construto dos instrumentos de avaliação docente, através da análise das qualidades psicométricas de um instrumento específico de avaliação do trabalho docente, respondido pelos estudantes de uma determinada Instituição de Ensino Superior (IES). Apesar do caráter local e específico do instrumento analisado, ele apresenta características presentes em boa parte dos instrumentos empregados nas mais diversas IES, de forma que a partir de alguns dos resultados deste estudo podemos inferir algumas implicações mais amplas e genéricas sobre a avaliação do trabalho docente por estudantes.

Apresentamos dois argumentos para questionar a validade de construto dos instrumentos de avaliação docente. Argumentamos que os instrumentos usuais de avaliação docente não distinguem, em sua mensuração, a percepção

¹ Contato:

E-mail: cristianogomes@ufmg.br

coletiva e a percepção individual dos estudantes a respeito do trabalho docente. Ao longo deste trabalho defenderemos a posição de que a percepção coletiva discente sobre o trabalho docente necessita de um tratamento metodológico técnico e consistente. O segundo argumento é que as formas usuais de instrumentos de avaliação docente não têm conseguido avaliar adequadamente a percepção dos estudantes a respeito do trabalho do professor, mas sim mensurar o julgamento do aluno a respeito desse trabalho. Argumentos e evidências serão apontados para suportar esse último argumento.

A avaliação do trabalho docente pelos discentes é importante se inserida em um programa de avaliação institucional que pretende desenvolver o ambiente acadêmico, promovendo o engajamento e melhorando as possibilidades de aprendizagem dos discentes. É nessa perspectiva que vemos uma aproximação entre a avaliação de docentes pelos discentes e a longa tradição de pesquisa sobre ambientes de aprendizagem. Nesta tradição de pesquisa os estudantes são considerados bons informantes sobre o ambiente escolar e de aprendizagem. Isso porque ao longo do processo de enculturação na escola, eles experienciam uma variedade de ambientes com tempo suficiente para formarem boas percepções dos aspectos mais estáveis dos ambientes, mesmo quando esses aspectos não são explícitos ou salientes (Frazer, 1998). Por essa razão, mesmo quando os professores são inconstantes em seu comportamento cotidiano, ainda assim os estudantes conseguem apreender as características distintivas e invariantes do ambiente. Por outro lado, o aluno não tem um acesso privilegiado ao conhecimento que detém e muitas vezes é incapaz de expressá-lo verbalmente, ou em registro escrito, de forma coerente e articulada, de forma que suas percepções a respeito do trabalho docente precisam ser coletadas de forma adequada e com instrumentos válidos e confiáveis.

As percepções dos estudantes sobre o ambiente escolar ou de aprendizagem podem ser pesquisadas a partir de duas perspectivas: uma que capta a visão pessoal e idiossincrática que cada estudante tem do ambiente e outra que capta a visão consensual, compartilhada socialmente a respeito do ambiente (Frazer, 1998). Ambas essas visões são informativas e importantes para os que planejam e desenvolvem ambientes de aprendizagem, pois as percepções dos estudantes e professores sobre o ambiente de aprendizagem expressam crenças sobre

as oportunidades de aprender e sobre o quanto tais oportunidades são limitadas pelo meio físico e social (Tobin & Frazer, 1998). O conhecimento dessas crenças propicia ao professor oportunidades para a ação (*affordances*), para projetar e modificar os ambientes de aprendizagem e, assim, interferir na história de aprendizagem do estudante. Ao agir sobre esses ambientes, o professor influencia o engajamento comportamental, cognitivo e emocional discente (Borges, Julio e Coelho, 2005), mesmo quando não consegue atingir seu objetivo (Moreira e Borges, 2006).

Do ponto de vista do desenvolvimento do ambiente acadêmico, numa perspectiva de avaliação institucional, o foco normalmente está em identificar aquelas oportunidades de ação que se apresentam à administração para promover o engajamento acadêmico e aumentar as possibilidades de aprendizagem dos discentes. É nessa perspectiva que conhecer a visão socialmente compartilhada pelos estudantes sobre o ambiente acadêmico torna-se relevante. Ela expressa o *ethos* da instituição e sua capacidade de promover o engajamento na perspectiva coletiva.

Os instrumentos de avaliação docente, em princípio, podem capturar duas visões, a pessoal e a consensual, coletiva. Um esforço de pesquisa importante consiste em tentar desenvolver instrumentos que capturem primordialmente uma das duas visões. Isso pode ser feito manipulando de forma sutil a redação das questões que compõem o instrumento (Frazer, 1998), ou ainda, modificando a unidade de análise (Cranton & Smith, 1990). Cranton e Smith (1990) mostram que um mesmo instrumento pode fornecer resultados diferentes quando os seus dados são analisados através de diferentes unidades de análise. Em seu trabalho empírico, eles usaram os dados da avaliação docente feita por 42407 estudantes de 2816 turmas distintas de cursos de comércio, utilizando um questionário já desenvolvido para avaliação do ensino por estudantes. Como unidades de análise eles usaram (1) as respostas individuais dos estudantes a cada um dos itens do questionário, (2) as médias por turma das respostas a cada item e o (3) desvio da resposta de cada estudante em relação à média da turma, por item do questionário. Para cada uma destas unidades de análise, a matriz de correlação inter-itens foi analisada usando a técnica de análise de fatores comuns, em que a comunalidade aparece na diagonal da matriz de correlação. Os autores obtiveram soluções com dois fatores em cada unidade de análise, que explicavam

grande parte da variância (88,6% no caso das respostas individuais, 91,8% no caso da média por turma e 87,3% no caso dos desvios das respostas individuais em relação à média da turma). Duas destas unidades de análise, a que considera cada resposta individual e a que considera o desvio de cada resposta individual em relação à média da turma, produziram estruturas com fatores congruentes entre si (coeficientes de congruência de 0,99 e 0,98 entre cada par de fatores). A unidade de análise que considera a média por turma produziu uma estrutura fatorial que não é congruente com as outras duas. Com base neste resultado, Cranton e Smith (1990) sustentam que o uso das médias por turma como unidade de análise elimina as diferenças individuais entre as percepções dos estudantes, incluindo as teorias implícitas e idiossincráticas, acentuando desta forma a visão coletiva e consensual.

Neste trabalho, como uma alternativa para capturar a visão consensual dos estudantes sobre o ambiente acadêmico, utilizamos como dados a serem analisados as médias dos resultados atribuídos pelos estudantes a cada docente, em cada um dos itens do instrumento de avaliação docente de uma IES. Consideramos a média por docente como a variável manifesta que expressa uma ou mais dimensões latentes que não temos acesso direto. Tais dimensões latentes sintetizam a percepção consensual e coletiva sobre o ambiente acadêmico. Elas expressam a percepção coletiva do *ethos* institucional. Através do procedimento descrito, acreditamos poder obter evidências sobre uma das questões fundamentais deste trabalho e que envolve a validade do instrumento de avaliação docente para a mensuração da percepção coletiva discente a respeito do trabalho do professor. Ao tratarmos os dados apontados, esperamos também analisar os itens do instrumento e discutir sua validade para capturar a percepção do trabalho docente pelo discente. Nossa hipótese é que o instrumento em questão, como a maioria dos instrumentos usuais neste tipo de avaliação, enfatiza o julgamento discente sobre o docente em detrimento dos aspectos perceptuais da natureza do trabalho docente.

Apresentamos na seção II deste trabalho uma breve revisão das pesquisas sobre a avaliação do trabalho docente pelos discentes. Esta revisão não pretende ser exaustiva, mas apontar um conjunto de lacunas e questões referentes às qualidades psicométricas dos instrumentos de avaliação docente e as implicações da utilização de

tais instrumentos nos ambientes educacionais do ensino superior. Os aspectos apontados nesta seção indicam, explícita ou implicitamente, através de argumentos e evidências, que os instrumentos internacionais usuais de avaliação docente não têm conseguido capturar a percepção dos estudantes sobre a natureza do trabalho docente. Ao contrário, têm incentivado um julgamento do aluno a respeito do que ocorre em sala de aula e do trabalho do professor. Esta seção também aponta argumentos e evidências de que a visão pessoal e idiossincrática dos estudantes tem sido preponderante, em função tanto da estrutura dos enunciados dos itens dos instrumentos de avaliação docente, quanto pelo modo como as IES vêm utilizando os resultados disponibilizados por instrumentos desse tipo.

AValiaÇÃO DOCENTE: UMA PERSPECTIVA INTERNACIONAL

Toda avaliação é uma ação definidora de valores. Ao avaliar o trabalho docente, uma Instituição de Ensino Superior (IES) determina, explícita ou implicitamente, um conjunto de critérios de qualidade para a comunidade universitária, composta pelos alunos, professores, gestores, entre outros. Esta ação definidora de valores traz em si um dos maiores problemas envolvidos com a avaliação docente e os seus instrumentos de mensuração. Martinson (2000) argumenta que o ambiente de ensino superior tem sido cerceado pelo que ele chama de imediatismo consumista. O autor sustenta que o imediatismo consumista é percebido na avaliação docente quando o professor é avaliado quase exclusivamente pelo prisma da satisfação do aluno. O foco central do seu argumento tenta evidenciar que várias IES vêm adotando a visão de que se o cliente (aluno) está satisfeito, as coisas devem estar bem. Em seu trabalho, Martinson (2000) discorre sobre uma série de argumentos e evidências sobre a falta de relação entre a opinião dos estudantes a respeito do trabalho docente e o desenvolvimento acadêmico e profissional discente. Buscando contrapor essa tendência, ele defende a idéia de que o trabalho docente deva ser avaliado pela sua efetividade em provocar e impulsionar o desenvolvimento dos estudantes, sugerindo o uso de formas de mensuração mais adequadas e capazes de identificar relações entre a qualidade de ensino e o crescimento da capacidade intelectual, atitudinal, ética e profissional dos alunos.

Martinson (2000) entende que a falta de uma análise mais adequada do trabalho docente não

se deve ao acaso, mas provém de uma postura fortemente mercadológica por parte de instituições que adotam uma política de valorização extrema do aluno como cliente. Ele alerta para o hiperconsumismo que avança para o campo da educação, através de uma perigosa associação entre compra e conhecimento, pagamento e transmissão de saber-fazer, embutida na visão do docente como prestador de serviços. Dentro do hiperconsumismo, o prestador de serviço (o professor) deve entregar ao cliente (o aluno) a obra pronta para ser consumida. E isto deve ser feito suprimindo da obra valores fundamentais como a necessidade de trabalhar séria e duramente, a motivação para ler, o estudar com afinco, etc.

Há outros trabalhos que reforçam a argumentação de Martinson (2000). Por exemplo, Crumbley e Fliedner (2002) apresentam evidências das quais se pode inferir que os argumentos de Martinson provavelmente espelham a realidade de muitas IES americanas e, especulamos, também de várias IES brasileiras. Evidências crescentes de supervalorização da opinião dos estudantes sobre o trabalho docente indicam que essa situação pode estar contribuindo para uma forte diminuição da qualidade do ensino, do rendimento acadêmico, do respeito aos professores, e para o enfraquecimento das instituições do ensino superior.

Estes autores discutem o que poderia ser considerado uma “epidemia” internacional: o problema da inflação das notas dadas aos alunos. Eles citam o exemplo de uma universidade americana na qual 80% de todas as notas dadas atualmente são conceitos A ou B, enquanto que nos anos de 1960 esse percentual era de 50%. Edwards (2000) analisou vários estudos que apresentam evidências a favor dessa realidade em um nível mais amplo. Uma meta-análise conduzida nos anos de 1990 e que usou dados relativos ao período de 1969 a 1993 para investigar o desempenho acadêmico de alunos de 4900 faculdades mostra que o número de conceitos A dados aos estudantes quadruplicou, embora o número de conceitos C tenha diminuído em 66%. Segundo Edwards (2000), poder-se-ia argumentar que o nível acadêmico dos estudantes aumentou ao longo dos anos. No entanto, ele exhibe evidências de que há um declínio em seus escores ao longo desse período nas provas do SAT (*Scholastic Aptitude Test*) e do ACT (*American College Training*), dois exames educacionais de larga escala do ensino superior americano. Deste fato, se infere que o inchaço das

notas não se deve à melhora no nível acadêmico dos estudantes, mas a outros fatores.

A inflação das notas oferece aos estudantes uma visão distorcida do seu desenvolvimento acadêmico e das habilidades adquiridas, impedindo um olhar mais apurado sobre a sua formação profissional (Edwards, 2000). Em seu trabalho, Edwards informa que várias instituições de ensino superior têm aumentado as notas atribuídas aos seus alunos como uma estratégia para impulsionar a auto-estima e motivar os estudantes. No entanto, segundo ele, essa estratégia tem efeito negativo a longo prazo.

Talvez o ponto mais problemático em relação à avaliação docente, até o presente momento, seja a fraca relação entre a avaliação do trabalho docente feita pelo aluno e o desenvolvimento de habilidades acadêmicas, profissionais, ou da capacidade de aprendizagem. Crumbley e Fliedner (2002) citam uma série de estudos que indicam haver uma fraca relação, quando há, entre o rendimento acadêmico e a avaliação docente. Yunker e Yunker (2003) endossam esta conclusão. Seu trabalho mostra que as correlações entre o desempenho acadêmico e a avaliação docente são, na verdade, negativas quando controladas por escores de provas nacionais de larga escala, como é o caso do ACT. Como boa parte dos estudos para analisar a correlação entre o desempenho acadêmico e a avaliação docente foca sua análise nas notas das próprias IES, o estudo de Yunker e Yunker (2003) alerta para a inadequação dessa estratégia metodológica e mostra a importância de serem utilizados instrumentos mais confiáveis de avaliação do desempenho acadêmico.

Os dados de Crumbley e Fliedner (2002) e Yunker e Yunker (2003) são corroborados pelas evidências de Pors (2001). Analisando mais de 10000 provas escolares de estudantes universitários entre 1988 e 1996, assim como 4200 questionários de avaliação docente, o autor não encontrou relações entre o desempenho acadêmico e a avaliação docente. Essa evidência é importante, pois a opinião dos estudantes sobre o trabalho docente não é afetada pelo produto ou pela melhora acadêmica (Pors, 2001). Tal resultado é importante por indicar que os estudantes podem não ser bons julgadores a respeito dos seus próprios processos de aprendizagem e das suas produções.

Outra evidência presente no estudo de Pors (2001) foi a verificação de que a única variável que se relaciona mais fortemente com a avaliação docente é a personalidade do professor, o que

permite aventar a hipótese de que a avaliação do trabalho docente pelo aluno é influenciada por relações de simpatia ou características pessoais dos docentes, e não por critérios de qualidade do ensino (Crumbley e Fliedner, 2002). Essa hipótese aparece também no estudo de Sally e Martin (2001). Os autores argumentam que uma parcela considerável das críticas feitas aos instrumentos de avaliação docente pelo estudante deve-se ao fato de que a avaliação docente não mensura níveis de qualidade de ensino, mas sim aspectos afetivos e relacionais articulados à personalidade do professor e a sua capacidade de ser popular e atrativo entre os estudantes. Estes autores realizaram uma regressão hierárquica analisando as variáveis tais como as notas dos alunos obtidas até metade do curso, a idade dos professores, o gênero e a idade dos estudantes, o tipo de matrícula feita pelo aluno para cursar uma disciplina, a habilidade acadêmica (medida pelo ACT), e o nível de extroversão dos professores percebidos pelos estudantes. Controlando todas essas variáveis, a extroversão foi a única variável capaz de explicar a variância do escore dos estudantes em relação ao instrumento de avaliação docente. Extroversão explicou quase 64% da variância do instrumento de avaliação docente, um resultado expressivo.

Os argumentos de Martinson (2000) e Crumbley e Fliedner (2002), entre outros, fazem concluir que há um forte incentivo para que os professores selecionem comportamentos e ações, inclusive disfuncionais, para ampliar o seu escore no instrumento de avaliação docente, de modo a obter vantagens salariais, promoção, etc, ou apenas para sobreviver no mercado de trabalho. Segundo Crumbley e Fliedner (2002), 39% dos administradores de faculdades entrevistados em sua pesquisa informaram conhecer professores que alteraram seu comportamento para melhorar seus escores no instrumento de avaliação docente.

Se as avaliações do trabalho docente são altamente influenciadas pelo contexto político e pedagógico das IES, elas também são muito influenciadas pelas concepções dos estudantes a respeito do trabalho docente. Há evidências importantes indicando que pouco se sabe a respeito do que mobiliza a percepção dos alunos a respeito do trabalho docente. Por exemplo, Langbein (1994) estudou a relação entre variáveis de características das disciplinas, características dos estudantes e características do professor e sua relação com a avaliação docente. As variáveis analisadas foram o tamanho da sala de aula, o gênero e o tempo de

docência do professor, os tipos de disciplinas, o tempo dedicado pelos estudantes fora da sala de aula com determinada disciplina, o desempenho acadêmico dos estudantes, o desempenho esperado pelo aluno em relação à disciplina, e a expectativa em relação à disciplina. Apenas 10% da variância dos escores dos estudantes no instrumento de avaliação docente são explicados por essas variáveis.

Há poucos estudos de meta-análise sobre o campo da avaliação docente, o que dificulta uma maior compreensão sobre as variáveis que influenciam a percepção dos estudantes. Além disso, algumas pesquisas (Husbands, 1998; Wachtel, 1998) incorporam em suas análises variáveis do mesmo tipo daquelas usadas no estudo de Langbein (1994), restringindo-se a analisar variáveis reconhecidamente limitadas para a explicação do fenômeno. Ao focar um conjunto de pesquisas e restringir o universo de artigos analisados nessa seção, procuramos apresentar discussões teóricas e evidências empíricas que pudessem salientar as questões fundamentais deste trabalho. Em síntese, os argumentos, evidências e discussões apontadas nesta seção reforçam a problemática da validade dos instrumentos de avaliação docente no que tange à sua capacidade de mensurar em nível distinto a percepção individual e coletiva dos estudantes, assim como também de mensurar aspectos perceptuais dos estudantes.

MÉTODO

Participantes

Nesse trabalho utilizamos os resultados médios da avaliação de um total de 638 professores de uma instituição de ensino superior de Belo Horizonte, Minas Gerais. Todos os alunos de todos cursos da instituição foram convidados a preencherem um instrumento de avaliação docente disponibilizado exclusivamente on-line. A instituição garantia aos alunos manter o sigilo total sobre suas respostas. A coleta e a agregação de dados foi terceirizada. A instituição não tinha acesso aos dados brutos, mas apenas a relatórios estatísticos preparados pela empresa contratada. Cerca de 30% dos estudantes da instituição participaram da avaliação docente de forma voluntária. Não há um relatório sobre a distribuição dos alunos participantes por cursos e períodos, bem como por gênero e idade. O resultado médio de cada docente, obtido no segundo semestre de 2005, foi disponibilizado aos pesquisadores pelo setor de

avaliação institucional da instituição. Não tivemos acesso às respostas individuais dos estudantes a cada instrumento. O resultado médio de cada docente é utilizado no trabalho para a análise do instrumento de avaliação docente.

Instrumento de Avaliação Docente e Procedimentos de Aplicação

O instrumento de avaliação docente aplicado no primeiro semestre de 2005 e analisado nesse artigo é um questionário que fazia parte de um conjunto de instrumentos da avaliação institucional de uma instituição de ensino superior de Belo Horizonte.

Basicamente, o instrumento de avaliação docente de 2005 busca mensurar três dimensões do trabalho docente: Didática e Avaliação, Postura Educadora, e Condução de Grupos. Os itens de um a cinco foram criados visando mensurar a dimensão Didática e Avaliação, os itens de 6 a 13 visavam a dimensão Postura Educadora e os dois últimos itens pretendiam acessar a dimensão Condução de Grupos. Os enunciados dos itens estão apresentados na Tabela 01.

Tratamento e Análise de Dados

Os dados tratados e analisados referem-se à pontuação média que cada professor recebeu de seus alunos nos 16 itens do instrumento de avaliação docente. Através desses dados foi utilizada a análise fatorial exploratória (AFE) para verificar se o instrumento de fato era capaz de mensurar as três dimensões visadas teoricamente. Dentre os procedimentos da AFE, foi selecionado o método dos eixos principais, a análise paralela por permutação e a rotação oblíqua oblímin. A preferência pelos referidos procedimentos da AFE sustenta-se em argumentos encontrados na literatura (Bentler, 1990; Bookstein, 1990; Buja & Eyuboglu, 1992; Carroll, 1993, 1995, 2003; Fisher, 1925; Gorsuch, 1974, 1990; Horn, 1965; Krazanowski & Kline, 1995; Lautenschlager, 1989; Loehlin, 1990; McArdle, 1990; Mulaik, 1990; Rozeboom, 1990; Schneeweiss, 1997; Thompson, 1994; Velicer & Jackson, 1990a, b; Widaman, 1990, 1993).

RESULTADOS

Através dos procedimentos da AFE utilizados, foi obtida uma solução de um fator, capaz de explicar 78,8% da variância comum dos

itens do instrumento. Os três primeiros autovalores extraídos foram 12,82; 0,55 e 0,49. A diferença entre o primeiro e o segundo autovalores é uma boa evidência da unidimensionalidade do instrumento e da adequação da solução de um único fator. Além disto, essa solução atende aos critérios clássicos de retenção de fatores autovalor maior do que um, *scree* teste, análise paralela por permutação. O *scree* teste visto na Figura 01 favorece a percepção e a interpretação de que um único fator captura a informação relevante do teste e que o instrumento é unidimensional. A Tabela 01 apresenta o enunciado, a comunalidade extraída e a carga fatorial de cada um dos 16 itens do instrumento.

Como vemos, todos os itens do instrumento apresentaram uma alta comunalidade extraída, indicando uma forte relação com o fator obtido. O item nove foi o que apresentou maior comunalidade (85%) e o item 13 a menor comunalidade (69%). Note-se que a menor comunalidade é ainda bastante acentuada, corroborando a forte relação entre os itens do instrumento e o fator encontrado. Ao examinar a confiabilidade da escala, encontramos um alto valor do *alpha* de Cronbach (0,98), indicando uma escala bastante confiável (Leontitsis & Pagge, 2007).

Dada a unidimensionalidade do instrumento e o alto percentual da variância explicada, quando necessário pode-se usar uma versão reduzida do instrumento, composta pelos sete melhores itens (um, dois, três, oito, nove, 11 e 15). A AFE dos sete itens selecionados indicou a presença de um único fator que explica 83,9% da variância comum desses itens, um resultado 5,1% a mais do que o obtido na análise do instrumento completo.

Os resultados apontados indicam que a diminuição do número de itens foi positiva e melhorou a explicação da variância pelo instrumento. O instrumento continuou apresentando uma escala altamente confiável (*alpha* de Cronbach de 0,97), mesmo tendo perdido mais da metade dos seus itens. Normalmente, a eliminação dessa quantidade de itens tende a baixar consideravelmente a confiabilidade do instrumento. No entanto, a alta comunalidade dos itens e sua forte carga no fator obtido não geraram decréscimo na precisão do instrumento, de modo que a versão reduzida com os sete itens mostrou-se a mais adequada para a mensuração da avaliação docente pelos estudantes.

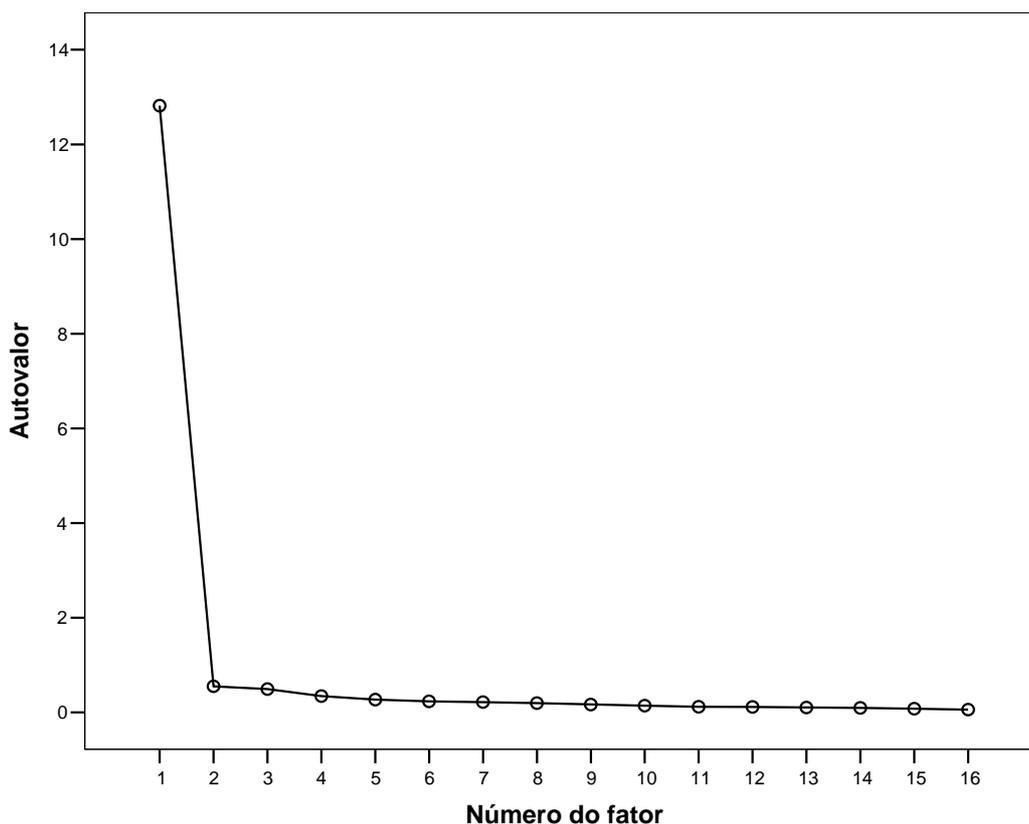


Figura 1. Scree teste relativo ao instrumento analisado

Tabela 1. Itens do Instrumento de Avaliação Docente, Comunalidade Extraída e Carga Fatorial

item	enunciado	comunalidade extraída	carga fatorial
1	trabalha o conteúdo com clareza, objetividade, organização e seqüência	0,83	0,91
2	correlaciona a disciplina ministrada com as demais disciplinas do curso	0,81	0,90
3	disponibiliza informações atualizadas sobre a disciplina	0,84	0,91
4	apresenta dicção/vocabulário e fluência verbal adequados	0,76	0,87
5	elabora avaliações condizentes com o conteúdo ministrado	0,78	0,88
6	disponibiliza e cumpre o Plano de Ensino da disciplina	0,78	0,88
7	planeja e prepara as aulas	0,75	0,87
8	está atento às dificuldades do aluno em sala e no conjunto do curso	0,84	0,92
9	incentiva a participação dos Alunos	0,85	0,92
10	esclarece dúvidas dos alunos sempre que solicitado	0,77	0,88
11	relaciona os aspectos teóricos da disciplina com seus aspectos práticos	0,84	0,92
12	mostra-se receptivo às críticas e sugestões dos alunos	0,76	0,87
13	incentiva as atividades extra-classe (pesquisa, eventos e projetos)	0,69	0,83
14	demonstra, valoriza e enfatiza os aspectos éticos da profissão	0,69	0,83
15	a liderança do professor produz harmonia no relacionamento com a classe	0,84	0,92
16	demonstra controle emocional em condições adversas	0,78	0,89

DISCUSSÃO

A solução obtida de um único fator pode ser considerada uma surpresa. Ela contrasta com uma visão bem difundida de multidimensionalidade do trabalho docente. As pesquisas sobre a formação e o desenvolvimento profissional dos professores apontam para a necessidade do professor dominar conhecimentos de diversos domínios, assim como desenvolver inúmeras competências profissionais, perspectivas e crenças sobre ensino e aprendizagem (Schulman, 1986). Se a visão multidimensional do trabalho docente é adequada era de se esperar que a percepção coletiva dos estudantes também fosse multidimensional. Algumas pesquisas sugerem haver base teórica e empírica para tratar a avaliação do trabalho docente feita pelos discentes como multidimensional (Marsh e Roche, 1997). Outras pesquisas apontam que os fatores encontrados se correlacionam fortemente entre si, corroborando a presença de um fator geral (d'Apollonia e Abrami, 1997; Greenwald & Gillmore, 1997, McKeachie, 1997).

Neste trabalho a avaliação do trabalho docente feita pelos discentes encontra-se articulada a um único fator, uma dimensão única. Assim, estamos diante de um instrumento com boas qualidades psicométricas, mas com problemas em relação à validade de construto. O que significa, afinal, essa dimensão?

Ao analisarmos a qualidade dos enunciados dos itens do instrumento estudado, podemos identificar que eles se atêm a obter uma medida a respeito do julgamento do aluno sobre o docente. O que falta a esse instrumento, como a vários outros do mesmo estilo, é a presença de itens com enunciados tais que ao respondê-los o aluno tenha de especificar como ele percebe o trabalho docente, identifica a maneira do professor de dar aula, percebe o modo como o professor transmite o conteúdo, entre outros aspectos do trabalho docente.

O “como é feito” não está presente nos instrumentos usuais de avaliação docente. Afinal, o que é “trabalhar o conteúdo com clareza, objetividade, organização e seqüência”, item presente no instrumento de avaliação docente avaliado neste trabalho? Será que, para um aluno, trabalhar o conteúdo com clareza é a mesma coisa que para outro aluno? O que significa para os alunos o professor ser objetivo? Quais comportamentos dos professores atraem os alunos e os satisfazem como critério de trabalhar o conteúdo com clareza e objetividade? “Correlacionar a

disciplina ministrada com as demais disciplinas do curso” também é um enunciado de item um tanto vago e não explica quais comportamentos dos professores se traduzem em uma percepção valorativa positiva dos estudantes. Este caráter vago e indefinido perpassa todos os itens: “disponibilizar informações atualizadas sobre a disciplina”, “estar atento às dificuldades do aluno em sala e no conjunto do curso”, “incentivar a participação dos alunos”, “relacionar os aspectos teóricos da disciplina com seus aspectos práticos”, “produzir harmonia no relacionamento com a classe”.

É substancialmente diferente um aluno dizer que para ele, aluno, o professor que está atento às dificuldades do aluno, em sala e no conjunto do curso, é aquele que marca horários extras para atender os alunos, que está disponível a qualquer momento para possíveis interações, que disponibiliza contatos na Internet etc, ou, ao contrário, é aquele que conversa com os alunos no horário da sala de aula, que solicita aos estudantes que façam perguntas e que tirem suas dúvidas durante a própria aula. Esse exemplo de ações e comportamentos diferentes implica em dizer que as ações que são valorizadas por um conjunto de estudantes podem ser radicalmente diferentes daquelas valorizadas por outro conjunto de alunos. Nesse sentido, não basta perguntar se a aula é “clara” ou não é. É necessário que o instrumento de avaliação possua itens capazes de capturar as percepções dos estudantes sobre o que seja uma aula “clara”. “Clareza” ou “objetividade” são pontos de vista, julgamentos que se embasam em um conjunto de percepções e sistemas de crenças, concepções e valores

Analisando os enunciados dos itens do instrumento interpretamos que o fator único obtido expressa um julgamento coletivo, socialmente compartilhado pelos discentes, sobre o trabalho docente. Parece que o instrumento captura um julgamento unidimensional justamente por ser composto por itens que mobilizam um julgamento sobre uma única dimensão: gostar ou não gostar da aula do docente, aspecto este que parece ser consensual e disponibilizado nitidamente em nível coletivo, dado que o instrumento explica fortemente a variância dessa dimensão.

Na seção II buscamos apresentar discussões teóricas e evidências que pudessem salientar as questões fundamentais deste trabalho. No que diz respeito à falta de validade dos instrumentos docentes em mensurar adequadamente a percepção dos estudantes sobre o trabalho

docente, os trabalhos revistos apontam que os instrumentos não conseguem abordar adequadamente a real natureza do trabalho docente. Apesar disso, esses resultados têm alterado a maneira de professores de dar aula para obter melhores avaliações por parte dos seus alunos. Alguns estudos revistos na seção II sugerem que os instrumentos internacionais de avaliação docente têm valorizado o julgamento do trabalho docente e têm sido boas ferramentas para a manutenção da exploração dos professores e para uma mercantilização da educação.

As características do instrumento específico aqui analisado são semelhantes a vários outros instrumentos nacionais e internacionais, e nosso questionamento sobre o instrumento específico pode ter amplitude e generalidade. Entendemos que, além das questões políticas e ideológicas apontadas na revisão realizada na seção II, há um problema técnico de validade de construto do instrumento. O que de fato os instrumentos de avaliação docente pelo discente pretendem medir?

Os resultados deste trabalho trazem algumas implicações para a construção e validação de instrumentos na área da avaliação docente. A primeira das implicações diz respeito ao fato de que o instrumento docente mostrou-se válido para identificar aspectos consensuais ou coletivos dos discentes a respeito dos docentes. Este fato é uma prova do conceito de que é possível construir instrumentos válidos para mensurar aspectos consensuais, socialmente compartilhados pelos discentes.

Os resultados encontrados não invalidam a idéia da avaliação do trabalho docente pelo discente. Entretanto, para ampliar a validade dessa avaliação, é necessário utilizar instrumentos de coleta de dados melhor desenvolvidos, o que significa melhorar a fundamentação teórica dos instrumentos e, ao mesmo tempo, utilizar metodologias bem conhecidas de desenvolvimento de instrumentos de coleta de dados. Nesse sentido, o estudo alude uma segunda implicação, e que envolve um vasto campo de pesquisa e investigação para o psicólogo na área da medida em avaliação docente. Esse campo apresenta algumas questões importantes que poderiam ser trabalhadas por psicólogos competentes na construção e validação de instrumentos e no uso de métodos estatísticos sofisticados.

REFERÊNCIAS

- Bentler, P. M. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, 25(1), 67-74.
- Bookstein, F. L. (1990). Least squares and latent variables. *Multivariate Behavioral Research*, 25 (1), 75-80.
- Borges, O.; Julio, J. M.. & Coelho, G. R. (2005). Efeitos de um ambiente de aprendizagem sobre o engajamento comportamental, o engajamento cognitivo e sobre a aprendizagem. Em R. Nardi & O. Borges (Orgs.) *Anais do V Encontro Nacional de Pesquisa em Educação de Ciências*, Bauru, SP : ABRAPEC. (CD-ROM)
- Buja, A. & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, 27(4), 509-540.
- Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies*. New York, NY: Cambridge University Press.
- Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research*, 30 (3), 429-452.
- Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: current evidence supports g and about ten broad factors. Em H. Nyborg (ed.), *The scientific study of general intelligence: tribute to Arthur R. Jensen*. (pp. 1-20). Oxford, UK: Elsevier Science/Pergamon Press.
- Cranton, P. & Smith, R. A. (1990). Reconsidering the unit of analysis: a model of student ratings of instruction. *Journal of Educational Psychology*, 82 (2), 207-212.
- Crumbly D. L. & Fliedner, E. (2002). Accounting administrators's perceptions of student evaluation of teaching (SET) information. *Quality Assurance in Education*, 10 (4), 213-222.
- d'Apollonia, S., & Abrami, P.C. (1997) Navigating student ratings of instruction. *American Psychologist*, 52, 1198-1208.
- Edwards, C. H. (2000). Grade inflation: the effects on educational quality and personal well being. *Education*, 120 (3), 538-546.
- Fisher, R. A. (1925). *Statistical methods for research workers*. s.e: s.c. Retirado em 01/02/2004, do World Wide Web: <http://psychclassics.yorku.ca>.
- Fraser, B.J. (1998). Science learning environments: assessment, effects and determinants. Em B.J Fraser & K.G Tobin (eds.). *International handbook of science education* . (pp. 527-564). Dordrecht, The Netherlands: Kluwer.

- Gorsuch, R. L. (1974). *Factor analysis*. Philadelphia: W. B. Saunders Company.
- Gorsuch, R. L. (1990). Common factor analysis versus component analysis: some well and little known facts. *Multivariate Behavioral Research*, 25(1), 33-39.
- Greenwald, A. G. & Gillmore, G. M. (1997) Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, 1209-1217.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Husbands, C. T. (1998). Implications for the assessment of the teaching competence of staff in higher education of some correlates of students' evaluations of different teaching styles. *Assessment and Evaluation in Higher Education*, 23 (2), 117-139.
- Krazanowski, W. J. & Kline, P. (1995). Cross-validation for choosing the number of important components in principal component analysis. *Multivariate Behavioral Research*, 30 (2), 149-165.
- Langbein, L. I. (1994). The validity of student evaluations of teaching. *Political Science & Politics*, 27 (3), 545.
- Lautenschlager, G. J. (1989). A comparison of alternatives to conducting Monte Carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research*, 24 (3), 365-395.
- Leontitsis, A. & Pagge, J. (2007). A simulation approach on Cronbach's alpha statistical significance. *Mathematics and Computers in Simulation*, 73 (5), 336-340.
- Loehlin, J. C. (1990). Component analysis versus common factor analysis: a case of disputed authorship. *Multivariate Behavioral Research*, 25 (1), 29-31.
- Marsh, H. W. & Roche, L. A. (1997) Making students' evaluations of teaching effectiveness effective: the critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187-1197.
- Martinson, D. L. (2000). Student evaluations of teaching and their short term validity. *Journalism & Mass Communication Educator*, 54 (4), 77-82.
- McArdle, J. J. (1990). Principles versus principals of structural factor analyses. *Multivariate Behavioral Research*, 25(1), 81-87.
- McKeachie, W. J. (1997) Student ratings: the validity of use. *American Psychologist*, 52, 1218-1225.
- Moreira, A. F. & Borges, O. (2006). Por dentro de uma sala de aula de física. *Educação e Pesquisa*. 32 (1), 157-174.
- Mulaik, S. A. (1990). Blurring the distinctions between component analysis and common factor analysis. *Multivariate Behavioral Research*, 25 (1), 53-59.
- Pors, N. O. (2001). Measuring students' performance and perceptions: empirical studies in different dimensions of quality assurance at a library school. *New Library World*, 102 (11/12), 429-435.
- Rozeboom, W. W. (1990). Whatever happened to broad perspective? *Multivariate Behavioral Research*, 25(1), 61-65.
- Sally, A. R. & Martin, D. J. (2001). Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis. *The Journal of Psychology*, 135 (3), 259-268.
- Schneeweiss, H. (1997). Factors and principal components in the near spherical case. *Multivariate Behavioral Research*, 32 (4), 375-401.
- Shulman, L. S. (1986) Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15 (2), 4-14.
- Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, 54 (4), 837-847.
- Tobin K. & Fraser, B.J. (1998). Qualitative and quantitative landscapes of classroom learning environment. Em B.J Fraser & K.G Tobin (eds.), *International handbook of science education* . (pp. 623-640). Dordrecht, The Netherlands: Kluwer.
- Velicer, W. F. & Jackson, D. N. (1990a). Component analysis versus common factor analysis: some further observations. *Multivariate Behavioral Research*, 25 (1), 97-114.
- Velicer, W. F. & Jackson, D. N. (1990b). Component analysis versus common factor analysis: some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25 (1), 1-28.
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: a brief review. *Assessment and Evaluation in Higher Education*, 23 (2), 191-211.

- Widaman, K. F. (1990). Bias in pattern loadings represented by common factor analysis and component analysis. *Multivariate Behavioral Research*, 25 (1), 85-95.
- Widaman, K. F. (1993). Common factor analysis versus principal component analysis? Differential bias in representing model parameters? *Multivariate Behavioral Research*, 28 (3), 263-311.
- Yunker, P. J., & Yunker, J. A. (2003). Are student evaluations of teaching valid? Evidence from an analytical business core course. *Journal of Education for Business*, 78 (6), 313-317.

Recebido em Junho de 2008

Reformulado em Agosto de 2008

Aceito em Outubro de 2008

SOBRE OS AUTORES:

Cristiano Mauro Assis Gomes: psicólogo, Doutor em Educação, é professor adjunto do Departamento de Psicologia da Universidade Federal de Minas Gerais e trabalha no campo da psicologia educacional e do desenvolvimento humano, avaliação psicológica e educacional.

Oto Neri Borges: físico, Doutor em Física, é professor associado do Departamento de Física da Universidade Federal de Minas Gerais, professor do Colégio Técnico de Minas Gerais e Coordenador do Programa de Pós-Graduação em Educação da Universidade Federal de Minas Gerais. Trabalha na área do ensino de ciências, cognição e avaliação educacional.