

CONSTRUÇÃO E VALIDAÇÃO DE UM INSTRUMENTO DE AVALIAÇÃO DO DESEMPENHO DOCENTE

Ilan Avrichir¹
Fernando Dewes²

Resumo

A avaliação do desempenho de professores universitários por seus alunos tem sido objeto de muitos estudos. Boa parte deles é no sentido de desenvolver critérios e formas de avaliação. Outra parte é direcionada para o questionamento e reflexão quanto aos problemas e possíveis benefícios da mesma. Por meio deste estudo desenvolveu-se uma escala destinada a avaliação do desempenho do professor. Suas propriedades psicométricas foram identificadas em duas fases. A primeira delas envolvendo 37 assertivas relacionadas ao desempenho do professor. Os resultados desta fase sugeriram a redução do número de assertivas, para possibilitar o emprego de ferramentas de estatística destinadas à identificação e análise das suas propriedades psicométricas. Na segunda fase do estudo foram encontrados índices satisfatórios de consistência e precisão. A análise fatorial sugeriu a presença de dois fatores, também com índices aceitáveis, indicando a validade do instrumento para o universo representado por sua amostra. Sugerem-se novos estudos para conferir maior amplitude de utilização do instrumento.

Palavras-chave: Avaliação de professores por alunos; análise fatorial; medidas psicométricas; validade de escalas.

Abstract

The evaluation of college professors by student has been studied extensively. Much effort has been invested in devising evaluation forms and criteria. Many studies have questioned and reflected about problems, biases and possible benefits accrued from evaluating professors. Through this study a scale was developed. Its psychometric properties were identified in two phases. In the first phase, 37 assertives related to professor's performance were developed. The results of this phase indicated the convenience of reducing the number of assertive. This was done using statistical techniques. In the second phase, in the application of the scale, satisfactory indexes of precision and consistency were found. The factorial analysis indicated the existence of two factors, also reliable, suggesting that the instrument has good validity. Further studies to confer bigger amplitude to the instrument are suggested.

Key Words: Professor evaluation by student; factorial analysis; psychometric measure; scale validity.

1. Introdução

Diversos estudos sugerem que a avaliação de professores universitários por estudantes pode contribuir significativamente para a melhoria da aprendizagem neste nível do ensino (Mckeachie e Kaplan, 1996; Cohen, 1981; Centra, 1993). No entanto, para que esta contribuição se torne possível é preciso que o processo de avaliação obedeça a determinadas características. Nos processos de avaliações que envolvem o preenchimento de um questionário por parte dos alunos, a natureza e redação das assertivas que compõem os instrumentos são críticas para a qualidade do processo. Este trabalho visa a contribuir para a melhoria dos processos de avaliação de professores por estudantes no Brasil da seguinte maneira:

1. traduz para a língua portuguesa e valida, com uma amostra de estudantes brasileiros, assertivas que se mostraram significativas em outros idiomas e culturas para a avaliação de professores;

¹ Escola Superior de Propaganda e Marketing - ESPM

² Escola Superior de Propaganda e Marketing – ESPM

2. torna público e disponível para uso por qualquer instituição de ensino um instrumento de avaliação, desenvolvido segundo técnicas recomendadas por autores especialistas no campo das medidas psicométricas (Nunnally, 1978; Pasquali 1999; Netemeyer, Bearden & Sharma 2003; Devellis, 2003);
3. relaciona um número grande de assertivas e descreve o processo de extração dos componentes principais a partir delas, de forma que qualquer pesquisador ou instituição pode gerar um instrumento validado para o seu contexto, se assim o quiser.

2. Desenvolvimento

A avaliação do desempenho de professores por estudantes é um dos temas mais frequentes na pesquisa sobre ensino e aprendizagem. Mckeachie e Kaplan (1996) sustentaram que já havia à época mais de 2000 artigos publicados sobre o assunto, sendo que mais de 1000 continham evidências trazidas pela pesquisa empírica. Centra (1993) informou que na base de dados do Ministério de Educação norte americano (ERIC), já haviam mais de 1500 artigos publicados sobre o tema.

O elevado número de estudos torna muito difícil qualquer tentativa de revisão detalhada da literatura. Por isto vamos, neste estudo, nos ater as meta-análises realizadas até aqui e nos artigos de autores reconhecidos que discutem especificamente as questões envolvidas na construção de instrumentos para avaliação de professores e, dentro deste campo, apenas à literatura sobre avaliação no ensino universitário. Intencionalmente deixamos de lado toda a pesquisa e controvérsias sobre a existência ou não de vieses nas avaliações feitas pelos estudantes, a discussão do impacto dos usos e abusos da avaliação, a questão da correlação entre satisfação dos alunos e aprendizagem, entre outras questões também relevantes ao tema. Focaremos somente trabalhos que nos orientaram na escolha do universo das assertivas das quais partimos para a construção do instrumento e justificaram a escolha das mesmas.

Feldman (1984) apud Centra (1993) fez uma síntese de trinta e um estudos que especificavam as características importantes do bom ensino. Este autor constatou uma grande similaridade entre os resultados dos estudos e entre as características apontadas por professores e alunos. Os fatores que apareceram com grande frequência e que o autor relaciona são: sensibilidade e preocupação com o nível e progresso da classe; preparação e organização do curso; conhecimento do assunto; entusiasmo; clareza; disponibilidade; justiça; imparcialidade na avaliação e qualidade das avaliações.

Sherman (1987) apud Lowman (2002) identificou cinco características, a saber: entusiasmo, clareza, preparação e organização, o aluno se sente estimulado e conhecimento. Lowman (2002) relaciona adjetivos que os estudantes usaram para descrever professores efetivos ou excelentes: entusiásticos conhecedores, inspiradores, bem humorados, interessantes, claros, organizados, envolventes, preparados, enérgicos, divertidos, estimulantes, criativos, apresentam bem, comunicativos, interessados, importam-se com os alunos, disponíveis, amistosos, acessíveis, respeitosos, compreensivos, encorajadores, desafiadores, justos, exigentes, pacientes, motivadores, dedicados e compromissados.

Feldman (1984) apud Centra (1993), após analisar vários estudos nos quais estudantes de graduação identificaram as características de professores excelentes, sustenta que elas podem ser agrupadas em doze categorias: despertar o interesse, entusiasmo, conhecimento do assunto, preparação e organização do curso,

clareza, eloquência, atenção ao progresso da classe, clareza dos objetivos do curso, relevância dos materiais, relevância dos materiais suplementares, carga e resultados percebidos pelos alunos.

Nas análises fatoriais dos instrumentos de avaliação também são observados resultados que apontam para características semelhantes. Fatores tais como: organização, planejamento ou estrutura do curso, interação professor-aluno, clareza da comunicação, dificuldade do curso, avaliação justa e auto-avaliação da aprendizagem estão presentes em muitos estudos fatoriais. A similaridade entre as qualidades identificadas por várias pesquisas como necessárias ao bom professor é, segundo Centra (1993), decorrente das pessoas normalmente associarem ensino com aulas expositivas. Este autor chama atenção para o fato de que caracterizar o bom ensino é mais complicado do que pode sugerir uma simples relação de atributos.

Conscientes de que a caracterização do bom ensino se constitui em matéria extremamente complexa, alguns autores recomendam cautela no emprego de critérios de avaliação do professor. Newport (1996) sugere, por exemplo, limitar os questionários a comportamentos dos professores que requerem pouca inferência e pouco julgamento sofisticado por parte dos alunos. Exemplos de comportamentos de baixa inferência são “o professor começou a aula pontualmente”, “o professor devolveu as provas corrigidas rapidamente” e “o plano de aula relacionava os objetivos da disciplina”. O mesmo autor recomenda a eliminação de itens que exigem muita inferência como “o professor foi habilidoso na observação das reações dos alunos” ou “o professor é um bom modelo dos comportamentos que advoga”.

Apesar das recomendações quanto ao nível de inferência requerido dos alunos pelos instrumentos de avaliação, Senior (1999) aponta que o uso de assertivas que requerem inferência elevada é, ainda, muito alto. Ele cita exemplos de vários instrumentos que examinou com perguntas sobre se “as tarefas pedidas pelo professor foram consistentes com os objetivos estabelecidos para o curso” ou se “o professor criou uma atmosfera conducente de aprendizado”. Perguntas formuladas desta maneira, além de induzirem interpretações pouco consistentes, limitam o valor do feedback como recurso para o aperfeiçoamento do desempenho do professor.

Segundo McKeachie e Kaplan (1996) há evidências de que os professores melhoram seus resultados em função de receberem feedback de avaliações feitas por alunos. Entretanto, a dimensão desta melhora depende do tipo de avaliação coletada e do uso feito desta informação. Tipicamente, itens de questionários que se referem a comportamentos específicos têm maior probabilidade de serem úteis do que itens gerais e abrangentes.

Sobre a necessidade das assertivas se referirem a comportamentos específicos Murray (1983) já havia comentado que há mais evolução do professor se itens relacionados a comportamentos forem usados em lugar de termos genéricos abstratos. Em vez de perguntar sobre a clareza, Murray (1985) sugere que seja perguntado se o professor “usou exemplos concretos” ou se “definiu os novos termos empregados”. Em vez de perguntar sobre a organização, deve-se perguntar se o professor “revê os tópicos da aula anterior” ou se “colocou o roteiro da matéria na lousa”. Como se observa, as recomendações acerca da formulação das assertivas são muito importantes para assegurar a consistência técnica dos instrumentos de avaliação e a utilidade dos seus resultados para a melhora do desempenho do professor avaliado.

Nossa busca de trabalhos acadêmicos sobre o tema da avaliação docente no Brasil se constituiu na varredura de três bases de dados disponíveis on-line. O SIBI-Dedalus, que indexa o material disponível em todas as bibliotecas da Universidade de São Paulo; o banco de teses da CAPES e a relação de artigos apresentados ao ENANPAD. Encontramos sete teses sobre avaliação do professor pelo aluno, mas apenas uma versando sobre questões técnicas envolvidas na construção de instrumentos. Esta amostra, por pequena que seja, sugere que os pesquisadores brasileiros, quando analisam o tema da avaliação docente tendem a se preocupar mais com o processo do que com o instrumento. Embora esta diferença de ênfase seja coerente com as importâncias relativas atribuídas pelos autores focalizados, a importância do instrumento não pode ser minimizada e ter questionários que dêem informações relevantes e utilizáveis pelos docentes e administradores é fundamental para a melhoria do desempenho dos professores.

Desconsiderar a importância do instrumento de avaliação poderá comprometer o alcance dos objetivos do sistema, levando a resultados pouco compensadores, tais como aqueles encontrados por Ambrósio (2001) que, analisando o processo de implantação da avaliação de desempenho dos professores na rede estadual de ensino de Minas Gerais, conclui que o processo é visto com descrédito pelos envolvidos e nada tem influenciado nas práticas pedagógicas dos docentes. Os objetivos pretendidos pelo sistema não estariam sendo, assim, alcançados. Parte disto poderia ser devida a inconsistências técnicas dos instrumentos utilizados para avaliação.

Moreira (1986) fez o que parece ser, ainda, o esforço mais longo e sistemático dado a conhecer publicamente no Brasil sobre o desenvolvimento de um instrumento de avaliação de professores por estudantes. Após rever a bibliografia disponível na época, submeteu a análise fatorial e discriminante o resultado de duas aplicações de um instrumento composto de 13 assertivas a amostras de cerca de 1300 alunos numa faculdade de Administração em São Paulo. As duas aplicações estiveram distantes entre si no tempo em cerca de três anos e o autor também comparou as avaliações dos cerca de 70 professores que foram avaliados nos dois momentos. O autor destaca, entre os resultados da pesquisa, o fato de não ter havido estabilidade temporal entre as avaliações dos professores (a avaliação do mesmo professor variou significativamente nas duas avaliações), o fato das dimensões fatoriais também não terem se mantido nem permitido uma interpretação clara e o fato de ter ocorrido o que ele chama de “efeito lógico” – os estudantes aparentemente pontuaram da mesma forma variáveis não relacionadas, independentemente da pessoa avaliada.

Embora o esforço e pioneirismo de Moreira ao realizar este estudo em 1986 não possam deixar de ser enaltecidos, o fato é que várias limitações transparecem hoje ao se analisar este estudo duas décadas após: o embasamento teórico das assertivas não foi suficientemente explicitado, a clareza das mesmas nem sempre atendeu aos critérios recomendados e a busca das dimensões fatoriais se deu sem que os fatores tenham sido claramente explicitados.

Entretanto, ao identificarmos estudos brasileiros voltados para analisar aspectos relativos ao processo da avaliação e dos seus resultados, independentemente das propriedades técnicas do instrumento, encontramos significativas contribuições. Carvalho (1990) analisou o processo de avaliação docente no Departamento de Eletrônica e Sistemas da Universidade Federal de Pernambuco usando métodos

predominantemente estatísticos e concluiu que as notas obtidas pelos discentes influenciam as avaliações que estes fazem dos seus docentes.

Pinho (2001) analisou, junto a professores de um curso de Ciências Contábeis de uma Instituição de Ensino Superior isolada, se e como o processo de avaliação do docente e no trabalho docente “sinalizam para a construção do professor reflexivo”. Conclui ser necessário criar foros específicos voltados para a discussão da atividade acadêmica em geral e avaliativa em particular e que contribuam para o amadurecimento do docente.

Ramos (1999) investigou o processo de avaliação do desempenho docente do Instituto de Química da Pontifícia Universidade Católica do Rio Grande do Sul. Trabalhou principalmente com a análise dos depoimentos de alunos acerca dos processos pedagógicos desenvolvidos no curso. Estes depoimentos foram analisados pelos professores do Instituto. Concluiu que a participação ativa dos professores em grupos organizados para analisar, compreender, refletir e teorizar sobre as suas práticas docentes contribui para o desenvolvimento profissional dos mesmos.

Rios (2004) realizou uma avaliação da avaliação docente de uma instituição de ensino superior usando análise documental e grupos focais com fundamentação no paradigma crítico de avaliação. Constata que o desenvolvimento da cultura de avaliação de docentes é fundamental para a reorientação curricular e enfatiza a importância do processo avaliativo do docente ser baseado em princípios de ética, solidariedade e humanização.

Em resumo, a revisão da literatura estrangeira revela uma série de prescrições para o desenvolvimento de instrumentos de medida relevantes e contributivos da melhoria de desempenho de professores. No entanto, a mesma revisão da literatura nacional não trouxe a tona nenhum esforço de validar as assertivas normalmente tidas como relevantes no exterior para o contexto brasileiro. Em contrapartida, observam-se importantes contribuições brasileiras para a compreensão do papel da avaliação, do impacto do seu processo nas percepções dos alunos e professores, bem como no estímulo a reflexões, de caráter filosófico ou prático, sobre avaliação do desempenho docente.

2.1 Metodologia e identificação de questões potenciais para instrumento

O desenvolvimento do instrumento obedeceu as seguintes etapas:

1. Identificação na literatura de assertivas que obedeceram as prescrições de Murray (1983; 1985) e Newport (1996);
2. Tradução e adaptação das assertivas;
3. Coleta de uma amostra de questionários numa Instituição de Ensino Superior em São Paulo para análise preliminar do instrumento (estudo piloto);
4. Coleta de amostras em três unidades da IES (São Paulo, Rio de Janeiro e Rio Grande do Sul) para o estudo de validação do instrumento.

Para a escolha das assertivas que comporiam o instrumento a ser submetido à análise das suas propriedades psicométricas partimos, principalmente, de Hativa, Barak e Simhi (2001). O estudo levado a efeito por estes autores se propôs a identificar os comportamentos e estratégias usadas por professores tidos como exemplares. Este estudo relacionou, a partir da literatura, trinta e quatro comportamentos e estratégias,

divididos em quatro categorias – organização da aula, clareza da aula, aulas interessante/envolventes e ambiente da sala de aula – como sendo as mais influentes no aprendizado e na satisfação dos alunos. Por ele ser recente e oferecer uma revisão cuidadosa da literatura e coerente com as recomendações para a formulação de assertivas, decidimos adotar os comportamentos identificados pelos autores como base para o nosso universo de assertivas.

O passo seguinte foi confrontar a relação inicial com a proposta por instrumentos consagrados de avaliação de professores por estudantes. Os principais destes instrumentos foram o Student Instructional Report (SRI), desenvolvido pelo Educational Testing Service da Universidade de Princeton; o Instructional Assessment System (IAS) desenvolvido pelo Educational Assessment Center da Universidade de Washington; o Instructional Development and Effectiveness Assessment (IDEA), desenvolvido pelo Centro para Avaliação e Educação Docente (Center for Faculty Evaluation and Development) da Universidade do Estado do Kansas e finalmente o Purdue's Cafeteria System.

O confronto entre estes instrumentos e a relação de itens que tínhamos mostrou um grande número de sobreposições, muitas vezes com as mesmas palavras. Identificamos, no entanto, duas grandes dimensões não cobertas. A primeira delas dizia respeito à percepção dos estudantes da justiça e adequação da avaliação do aprendizado feita pelo professor. A segunda dizia respeito a questões que avaliassem globalmente a satisfação dos estudantes com o curso. O instrumento que resultou do confronto apresentou, então, trinta e sete assertivas relativas a comportamentos de professores, sendo três delas referentes à satisfação geral do aluno com a disciplina.

Esta relação foi discutida com professores das diversas unidades da Instituição de Ensino Superior - IES onde foi realizado o estudo, que a criticaram e propuseram alterações semânticas em algumas assertivas. Com o propósito de identificar preliminarmente algumas propriedades psicométricas deste instrumento foi realizado, então, um estudo piloto, cujos resultados guiaram os passos subseqüentes na busca do refinamento do mesmo.

2.2 Instrumento

As 37 assertivas que resultaram do levantamento inicial foram colocadas no formato de uma escala de Likert com cinco pontos ou graus. Algumas assertivas foram redigidas na direção negativa, como tentativa para diminuir o efeito de halo; efeito pelo qual o respondente generaliza uma resposta ao conjunto do questionário e responde várias perguntas sem refletir ou, mesmo, quase sem lê-las. O instrumento inicial está apresentado no Anexo 1.

2.3 Procedimentos

O instrumento foi aplicado em uma amostra de 224 estudantes da unidade da IES localizada em São Paulo, distribuídos em 9 classes. Estas classes foram selecionadas de forma a ter turmas dos vários semestres e cursos, de disciplinas da área de Humanas e Exatas e professores que costumavam ser avaliados pelos alunos de forma positiva, isto é, com médias mais elevadas e outros apresentando médias mais baixas. Os questionários foram aplicados por um mesmo professor em todas as classes. Foi explicado aos alunos o

objetivo do preenchimento e a importância da contribuição deles. Os questionários preenchidos foram tabulados e alimentados numa base de dados SPSS e os seus resultados submetidos à análise estatística.

2.4 Resultados

O instrumento, na sua concepção inicial, mostrou-se inadequado para ter seus itens submetidos à análise fatorial. O elevado número de itens carregando em um único fator e a não aceitação da rotação, como forma de buscar um melhor equilíbrio fatorial, sugeriram a unidimensionalidade do instrumento. Esta particularidade não justificaria, portanto, a adoção de uma gama elevada de itens, a menos que se agrupassem em distintas dimensões.

Desta forma, os resultados iniciais induziram a realização de alguns ensaios envolvendo itens com diferentes conteúdos e quantidades. Itens que, por exemplo, apresentavam significados muito próximos ou semelhantes eram excluídos de cada ensaio, reduzindo-se assim a superposição de conteúdos. Também foram excluídos itens formulados na direção negativa, pois, através dos ensaios, observou-se que a presença dos mesmos distorcia o comportamento das médias.

Tais ensaios proporcionaram diversas rodadas de análise fatorial, até chegar naquela que apresentou indicadores da existência das características mais desejáveis num instrumento sem prejuízo, naturalmente, da abrangência dos conceitos teóricos envolvidos. A seqüência de ensaios foi dada por encerrada quando se chegou a um conjunto de 16 itens que apresentaram os seguintes indicadores estatísticos:

a. Consistência Interna

Com base no coeficiente Alpha de Cronbach, o conjunto composto por 16 itens apresentou uma consistência interna aceitável ($\alpha = 0,8790$). Quanto maior o coeficiente Alpha, maior a consistência interna da escala. Para Nunnally (1978), um coeficiente Alpha aceitável de uma escala deveria ser de, no mínimo, 0,70.

Para sondar o grau de precisão do conjunto utilizou-se o método das duas metades (split-half). Esse método agrupa os itens em dois conjuntos. Quanto maior a correlação entre os conjuntos, maior a precisão. A correlação entre as duas metades, segundo o índice de Guttman, foi de 0,8730, o que indica a validade de precisão do instrumento, confirmando também a sua consistência interna.

b. Análise Fatorial

Para determinar a adequação da utilização do método de análise fatorial, extraiu-se inicialmente o índice KMO (Kaiser-Meyer-Okin). O valor obtido foi de 0,896, sugerindo a viabilidade do método para análise dos dados coletados, conforme Kaiser (1974). Com propósito semelhante, efetuou-se o teste de Bartlett e obteve-se o valor de 1278,894, confirmando a pertinência do modelo fatorial (NORUSIS, 1994).

A média das comunalidades (quantidade de variância que um item ou variável compartilha com todos os demais itens) foi de 0,542. Com exceção de um único item, todos os demais apresentaram comunalidade final superior a 0,40.

A extração da carga fatorial inicial sugeriu a possibilidade de três fatores, tendo o de menor carga um peso de 8,18% da variância total explicada. Com três fatores já foi possível explicar mais de 54% da variância total existente no conjunto das variáveis presentes na concepção do instrumento.

Entretanto, o desequilíbrio entre as cargas dos diferentes fatores recomendava buscar uma distribuição relativamente mais equilibrada das variâncias explicadas pelos fatores. Através da rotação ortogonal Varimax foi possível, então, encontrar uma melhor distribuição dos percentuais das variâncias explicadas por cada fator, como pode ser verificada na tabela abaixo, que apresenta a situação após a extração inicial dos fatores e após a rotação dos mesmos.

Tabela 1 - Resultados da análise fatorial inicial e após a rotação

Fator	Extração inicial			Após a rotação dos fatores		
	Autovalor	% variância	% cumulativo	Autovalor	% variância	% cumulativo
1	5,916	36,978	36,978	3,491	21,817	21,817
2	1,456	9,102	46,080	2,982	18,638	40,455
3	1,309	8,179	54,259	2,209	13,804	54,259

b.1 Composição dos fatores

As Tabelas 2, 3 e 4 apresentam as informações essenciais sobre os fatores e os itens agrupados conforme os resultados da análise fatorial rotacionada. Nelas se observa que todos os itens carregam, em algum fator, uma magnitude de carga superior a 0,40, sugerindo serem eles representantes úteis dos fatores encontrados (PASQUALI, 1999). Naturalmente, quanto maior a carga fatorial, melhor o item e, neste sentido, muitos deles apresentam cargas fatoriais respeitáveis.

Tabela 2 - Fator 1 - Interesse despertado pelo curso

Item	Descrição	Dimensão original	Carga fatorial
21	O professor usou estratégias pedagógicas variadas (casos, dinâmicas, vídeos, etc.)	Interesse	,474
24	O professor me defrontou com desafios estimulantes	Interesse	,745
26	Os assuntos abordados foram de interesse para o meu desenvolvimento pessoal ou profissional	Interesse	,753
18	O professor fez apresentações de forma entusiástica e dinâmica	Interesse	,688
30	O professor ofereceu feedback estimulante aos alunos	Clima	,496
27	Os trabalhos exigidos foram estimulantes e interessantes	Interesse	,728
36	Este professor é um dos melhores que já tive	Geral	,698

Tabela 3 - Fator 2 - Clima na sala de aula

Item	Descrição	Dimensão original	Carga fatorial
28	O professor demonstra interesse pelos alunos	Clima	,747
2	O professor explicou a relação das aulas com as aulas anteriores	Organização	,458
9	O professor respondeu com cuidado e precisão às perguntas dos alunos	Clareza	,767
31	O professor foi acessível	Clima	,598

Tabela 4 - Fator 3 – Avaliação e clareza

Item	Descrição	Dimensão original	Carga fatorial
7	O professor fez perguntas para verificar se os alunos estavam entendendo	Clareza	,437
10	O professor enfatizou os pontos importantes	Clareza	,587
32	O professor explicou como os alunos seriam avaliados	Avaliação	,670
25	O professor usou humor	Interesse	,412
33	As provas refletiram os pontos importantes do curso	Avaliação	,579

Ao examinar as tabelas acima, percebe-se que a análise fatorial constituiu grupos de itens com identidade de conteúdo com as dimensões preconizadas à priori. No fator 1 estão reunidos itens predominantemente referentes ao interesse que o curso desperta no estudante.. No segundo fator, dois dos itens dizem respeito ao relacionamento professor-estudante. O item “o professor respondeu com cuidado e precisão as perguntas dos alunos”, apesar de ter sido classificado de início como pertencente à dimensão clareza, pode ter sido facilmente interpretado como relacionado ao interesse do professor pelo estudante. No terceiro fator apareceram duas assertivas relativas à justiça na avaliação e duas ligadas à clareza. Com exceção do item relativo ao uso do humor, os demais parecem ligados à como o professor estrutura o curso, ao que em teoria de liderança se costuma chamar “dimensão tarefa”.

2.5. Estudo de validação

Os resultados obtidos através do estudo piloto mostraram índices animadores quanto ao potencial do instrumento, sugerindo a viabilidade do prosseguimento do estudo, partindo de uma amostra mais ampla, de modo a permitir uma análise mais refinada das propriedades psicométricas do instrumento.

2.6 Método

A partir dos itens extraídos do estudo piloto foi composto o novo instrumento. Os 16 itens foram arranjados em uma escala do tipo Likert, desta feita com sete graus. Além disto não se usaram mais itens descritos na direção negativa, pelas razões já expostas anteriormente.

O instrumento foi aplicado nas três unidades da IES (São Paulo, Rio de Janeiro Porto Alegre), ficando a amostra inicial constituída por 518 alunos e vinte professores avaliados. A unidade São Paulo participou com 239 alunos e sete professores, Rio de Janeiro com 151 alunos e seis professores e Porto Alegre com 128 alunos e sete professores. A composição da amostra atendeu ao critério de disponibilidade, visto que muitas turmas estavam envolvidas, à época da aplicação, em outras atividades que impossibilitavam sua participação.

A aplicação em todas as turmas, de cada unidade da IES, foi feita por uma única pessoa, na própria sala de aula de cada turma.

2.7 Resultados

Os resultados preliminares do estudo de validação apresentaram índices estatísticos mais favoráveis no que diz respeito à consistência interna do instrumento, com relação àqueles obtidos no estudo piloto. O índice de 0,9190, correspondente ao coeficiente Alpha, superou o anterior, que foi de 0,8790. Contudo as estatísticas

referentes à Item-total revelaram que, se o item 3 fosse eliminado, obter-se-ia um índice de consistência interna de 0,9417. Além disto, a extração das médias das comunalidades (quantidade de variância que um item ou variável compartilha com todos os demais itens) indicou no item 3 uma comunalidade final de apenas 0,217, bem inferior àquelas apresentadas nos demais itens.

Estes índices estimularam a realização de novos testes estatísticos excluindo o item 3. Nesta condição extraiu-se, então, o índice Alpha que se revelou mais aceitável em relação aos índices anteriores, alcançando o coeficiente de 0,9418.

O grau de precisão do conjunto, aferido através do método das duas metades (split-half), também se mostrou superior ao anterior, alcançando 0,9341 no índice de Guttman.

Os indicadores fornecidos pela análise fatorial também se mostraram superiores aos do estudo anterior. A tabela das comunalidades, apresentada abaixo, revela que, à exceção do item 7, todos os demais itens estão acima de 0,50.

Tabela 6 - Estudo de Validação- Comunalidades

Item	Inicial	Extração
1	1,000	,532
2	1,000	,658
4	1,000	,711
5	1,000	,587
6	1,000	,610
7	1,000	,462
8	1,000	,662
9	1,000	,687
10	1,000	,656
11	1,000	,664
12	1,000	,700
13	1,000	,511
14	1,000	,544
15	1,000	,677
16	1,000	,749

A análise fatorial rotacionada revelou a presença de dois fatores, como se pode observar na tabela a seguir:

Tabela 7 - Estudo de validação - Resultados da análise fatorial inicial e após a rotação

Fator	Extração inicial			Após a rotação dos fatores		
	Autovalor	% variância	% cumulativo	Autovalor	% variância	% cumulativo
1	8,352	55,681	55,681	4,752	31,679	31,679
2	1,057	7,048	62,729	4,657	31,050	62,729

De um modo semelhante aos índices anteriores, percebe-se nesta tabela que o percentual acumulado da variância explicada pelos fatores foi superior ao estudo anterior, alcançando, desta feita, 62,729 %. Ademais, verifica-se um equilíbrio na participação de cada fator, promovido pela melhor distribuição dos percentuais das variâncias explicadas (31,679 e 31, 050).

Composição final dos fatores e análise dos mesmos

As tabelas abaixo especificam os itens agrupados pela análise fatorial e suas respectivas cargas:

Tabela 8 - Fator 1 – Interesse e desafio

Item	Descrição	Carga fatorial
8	Os assuntos abordados foram de interesse para o meu desenvolvimento pessoal ou profissional	0,800
4	O professor me defrontou com desafios estimulantes	0,786
11	Os trabalhos exigidos foram interessantes	0,738
9	O professor fez apresentações de forma entusiástica	0,697
16	Este professor é um dos melhores que já tive	0,692
1	O professor usou estratégias pedagógicas variadas (casos, dinâmicas, vídeos, etc.)	0,676
5	O professor explicou a relação das aulas com as anteriores	0,588
10	O professor ofereceu feedback construtivo aos alunos	0,584

Tabela 9 - Fator 2 – Relacionamento e avaliação

Item	Descrição	Carga fatorial
15	O professor foi acessível	0,785
12	O professor respondeu com cuidado às perguntas dos alunos	0,778
2	O professor demonstrou interesse pelos alunos	0,701
13	O professor usou humor	0,658
7	O professor explicou como os alunos seriam avaliados	0,655
14	As provas refletiram os pontos importantes do curso	0,637
6	O professor enfatizou os pontos importantes	0,583

Os conteúdos dos itens agrupados no fator 1 convergem para duas direções. A primeira delas parece referir-se ao grau de interesse que o professor é capaz de despertar no aluno através dos assuntos apresentados. A segunda diz respeito à forma como apresenta os conteúdos da matéria. Uma questão crucial se coloca aqui: Como tornar a matéria mais atraente ou interessante para o aluno?

Uma pista pode ser identificada para responder, ainda que preliminarmente, esta questão. Por meio das correlações entre os itens (apresentadas no anexo 2) é possível verificar que o maior índice de correlação é encontrado entre os itens 16 e 4 (0,6848). O conteúdo do item 16 pode ser considerado o mais genérico de todos, pois envolve uma avaliação geral do desempenho do professor, sem considerar uma dimensão específica. O item 4, por sua vez, está ligado à uma dimensão específica, a do desafio. Desafiar o aluno parece ser um caminho compensador para mantê-lo interessado no curso. Mas como tornar um desafio estimulante? É outra questão que surge. Para respondê-la, ainda que de uma forma meramente especulativa, podemos voltar às correlações entre os itens. Através das correlações se verifica que o entusiasmo com que o professor apresenta os conteúdos da matéria e o feedback oferecido ao aluno estão associados, de uma forma expressiva, na resposta à esta questão. As elevadas correlações dos itens 9 e 10 com o item 16 autorizam esta especulação.

Já no segundo fator aparecem itens de conteúdo referente ao atendimento ao aluno, como pessoa. Atenção, cuidado, interesse e justiça na avaliação, constituem as idéias centrais dos itens. Consideração pessoal, expressa pelo interesse do professor com cada aluno, respondendo suas perguntas com cuidado e mostrando-se disponível através de gestos que incluem também o bom humor, parecem também estar associados ao bom desempenho do professor, ao menos segundo a percepção do aluno.

3 Conclusões

O trabalho feito sugere, fortemente, que um instrumento com 15 itens, conforme apresentado no anexo 3, sobre comportamentos específicos e um geral pode avaliar a satisfação dos alunos. No entanto, este estudo não testou a estabilidade temporal do instrumento e se limitou a validá-lo no âmbito de uma instituição de ensino. Embora, neste âmbito ele tenha se mostrado satisfatório em 3 cidades diferentes, novas validações em outras IESs, localidades e cursos são necessárias antes de podermos afirmar que este instrumento pode ter sua aplicação generalizada. Também nos compete alertar para o fato de que cautela deve ser exercida antes de assumirmos que o instrumento possa ser válido em outros níveis de ensino. Como nos lembra Murray (1985) a avaliação de professores que ensinavam em graduação e pós-graduação tem baixa correlação. Ou seja, professores que se saem bem em um nível não necessariamente se saem bem quando ensinam em outro nível. Isto sugere que não devemos assumir que as assertivas relevantes sejam as mesmas, sem uma prévia verificação empírica.

Referências

- AMBRÓSIO, I.R. *Avaliação de desempenho ou desempenho da avaliação dos docentes*. Dissertação de mestrado apresentada à Universidade Federal de Juiz de Fora, 2001.
- CARVALHO, E. *Desempenho docente: o caso do Departamento de Eletrônica e Sistemas da Universidade Federal de Pernambuco*. Dissertação de Mestrado apresentada a Faculdade de Administração da Universidade Federal de Santa Catarina, 1990.
- COHEN, P.A. Student rating of instruction and student achievement: a meta-analysis of multisection validity studies. *Review of Educational Research* n. 51, p. 281-309, 1981.
- CENTRA, J.A. *Reflective faculty Evaluation: enhancing teaching and determining faculty effectiveness* NY, Jossey Bass. 1993.
- DEVELLIS, R.F. *Scale Development: theory and applications*. Thousand Oaks, California Sage. 2003
- HATIVA, A. BARAK, F. e SIMHI, J. Exemplary university teachers: knowledge and beliefs regarding effective teaching dimensions and strategies. *The journal of higher education* Vol 72, n. 6, nov-dez 2001 699-729, 2001.
- KAISER, H.F. An index of factorial simplicity. *Psychometrika*, v. 39, p. 31-36, 1974
- LOWMAN, J. *Dominando as técnicas do ensino*. São Paulo, Atlas, 2002.
- MCKEACHIE W. J., KAPLAN, M. Persistent problems in evaluating college teaching. Disponível em http://cedar.olemiss.edu/depts/vc_academic_affairs/problems.html, 1996.
- MOREIRA, D.A. *Avaliação do professor universitário pelo aluno: possibilidades e limitações*. Tese de doutoramento apresentada à Faculdade de educação da Universidade de São Paulo. 1986.
- MURRAY, H.G. Low-inference classroom teaching behaviors and student ratings of teaching effectiveness in *Journal of educational psychology* 75, 1983, p. 138-49.

Revista Eletrônica de Ciência Administrativa (RECADM) - ISSN 1677-7387
Faculdade Cenecista de Campo Largo - Coordenação do Curso de Administração

v. 5, n. 2, nov./2006 - <http://revistas.facecla.com.br/index.php/recadm/index>

MURRAY, H.G. *Classroom teaching behaviors related to college teaching effectiveness*. In Donald, J.G; Sullivan A.M. (editors) *Using research to improve teaching*. New Directions in Teaching and Learning, n23, San Francisco; Jossey Bass, p.21-34, 1985.

NETEMEYER, R.G; BEARDEN, W.O.; SHARMA, S. *Scaling procedures: issues and applications*. Thousand Oaks, California Sage, 2003.

NEWPORT, J. F. Rating Teaching in the USA: Probing the qualifications of student raters and novice teachers. *Assesment & Evaluation in Higher Education*, 21, n1, p 17-21, Mar, 1996.

NORUSIS, M.J. *SPSS professional statistic 6,1*. Chicago: SPSS, 1994.

NUNNALLY, J. C. *Psychometric theory*. 2ed. New York: McGrawHill, 1978.

PARDO, A.; RUIZ, M. A. *SPSS 11: Guia para el análisis de datos*. Madrid: McGraw Hill, 2003.

PASQUALI, L. *Instrumentos psicológicos: Manual prático de elaboração*. Brasília: LabPAM/ IBAPP, 1999.

PINHO, L. F. *Trabalho docente e práticas avaliativas numa instituição de ensino superior isolada: expectativas e possibilidades*. Dissertação de mestrado apresentada à Faculdade de Educação da Universidade Metodista de Piracicaba, 2001.

RAMOS, M. G. *Avaliação do desempenho docente numa perspectiva qualitativa: contribuição para o desenvolvimento profissional de professores no ensino superior*. Tese de Doutorado apresentada à Faculdade de Educação da Pontifícia Universidade Católica do Rio Grande do Sul 1999.

RIOS, M.P.G. *Meta-avaliação de docentes no ensino superior*. Tese de Doutorado apresentada à Faculdade de Educação da Pontifícia Universidade Católica de São Paulo. 2004.

SENIOR, B. A. *Student Teaching Evaluations: Options and Concerns*. *ASC Proceedings of the 35th Annual Conference* California Polytechnic State University - San Luis April 7 - 10, 1999. pp 251 – 260.

SMITH, W. *A review of the research into teaching/styles behaviors' impact on students's cognitive outcomes*. Paper presented at the annual meeting of the southwest educational research association Austin, TX, January 23-25, 1997.

Anexo 1 - Assertivas e dimensões preconizadas

1	O professor explicou como os alunos seriam avaliados	Avaliação
2	As provas refletiram os pontos importantes do curso	Avaliação
3	O professor fez perguntas para verificar se os alunos estavam entendendo	Clareza
4	O professor desencorajou os alunos a fazerem perguntas*	Clareza
5	O professor respondeu com cuidado as perguntas dos alunos	Clareza
6	O professor enfatizou os pontos importantes	Clareza
7	O professor escreveu na lousa de forma legível	Clareza
8	O professor deu poucos exemplos e ilustrações esclarecedoras*	Clareza
9	O professor repetiu e elaborou os pontos difíceis	Clareza
10	O professor explicou os termos difíceis ou não familiares	Clareza
11	O professor cometeu enganos e erros*	Clareza
12	O professor falou de forma dificilmente audível*	Clareza
13	As diretrizes para a execução dos trabalhos foram claras e precisas	Clareza
14	O professor demonstra interesse pelos alunos	Clima
15	O professor comportou-se de forma desrespeitosa com os alunos*	Clima
16	O professor foi acessível.	Clima
17	As notas foram dadas de forma injusta e parcial	Geral
18	Este curso é um dos melhores que já fiz	Geral
19	Este professor é um dos melhores que já tive	Geral
20	Eu gostaria de ter um outro curso com este professor	Geral
21	O professor usou analogias, metáforas e anedotas	Interesse
22	O professor explicou a origem histórica e filosófica dos conceitos	Interesse
23	O professor usou estratégias pedagógicas variadas (casos, dinâmicas, vídeos, etc...)	Interesse
24	O professor referiu-se à sua própria vida ou experiências profissionais	Interesse
25	O professor integrou ao curso temas vistos em outras disciplinas	Interesse
26	O professor me defrontou com desafios estimulantes	Interesse
27	O professor usou humor	Interesse
28	Os assuntos abordados foram de interesse para o meu desenvolvimento pessoal ou profissional	Interesse
29	O professor fez apresentações de forma entusiástica e dinâmica	Interesse
30	O professor ofereceu feedback estimulante aos alunos	Interesse
31	Os trabalhos exigidos foram estimulantes e interessantes	Interesse
32	O professor explicou a relação das aulas com o programa da disciplina	Organização
33	O professor explicou a relação das aulas com as aulas anteriores	Organização
34	O professor apresentou o roteiro das aulas no início das aulas	Organização
35	O professor fez um resumo da aula ao final da aula	Organização
36	O professor dividiu a aula em tópico ou partes	Organização
37	O professor deixou claro quando terminava um tópico e começava um outro	Organização

*Perguntas apresentadas de forma invertida para diminuir a tendência dos respondentes se fixarem num valor e responderem todas as perguntas com ele.

Anexo 2 – Itens do instrumento final

O professor demonstra interesse pelos alunos	1	2	3	4	5	6	7
O professor explicou a relação das aulas com as aulas anteriores	1	2	3	4	5	6	7
O professor respondeu com cuidado às perguntas dos alunos	1	2	3	4	5	6	7
O professor usou estratégias didáticas variadas (casos, dinâmicas, vídeos, etc...)	1	2	3	4	5	6	7
O professor enfatizou os pontos importantes	1	2	3	4	5	6	7
O professor me defrontou com desafios estimulantes	1	2	3	4	5	6	7
O professor explicou como os alunos seriam avaliados	1	2	3	4	5	6	7
O professor usou humor	1	2	3	4	5	6	7
Os assuntos abordados foram de interesse para o meu desenvolvimento pessoal ou profissional	1	2	3	4	5	6	7
O professor fez apresentações de forma entusiástica	1	2	3	4	5	6	7
O professor ofereceu feedback construtivo aos alunos	1	2	3	4	5	6	7
O professor foi acessível.	1	2	3	4	5	6	7
Os trabalhos exigidos foram interessantes	1	2	3	4	5	6	7
As provas refletiram os pontos importantes do curso	1	2	3	4	5	6	7
Este é um dos melhores professores que já tive	1	2	3	4	5	6	7

Anexo 3 – Correlações entre assertivas – Instrumento final

Correlation Matrix

	Q1	Q2	Q4	Q5	Q6
Q1	1,0000				
Q2	,5606	1,0000			
Q4	,5399	,5671	1,0000		
Q5	,4487	,5583	,6168	1,0000	
Q6	,4286	,5536	,5853	,7150	1,0000
Q7	,3181	,4021	,3386	,4326	,4981
Q8	,4409	,3762	,6025	,4810	,4886
Q9	,5495	,6020	,6296	,5545	,5907
Q10	,5518	,5996	,6001	,6257	,5960
Q11	,5885	,4926	,5967	,4955	,5015
Q12	,4484	,6394	,4473	,5250	,5649
Q13	,3154	,5400	,4192	,4079	,4385
Q14	,3764	,4885	,5045	,5123	,5783
Q15	,4154	,6845	,4610	,4503	,5104
Q16	,5429	,6394	,6846	,6528	,6236

Q7 Q8 Q9 Q10 Q11

Q7	1,0000				
Q8	,3247	1,0000			
Q9	,3452	,5860	1,0000		
Q10	,4720	,4654	,6326	1,0000	
Q11	,4206	,5900	,6531	,6252	1,0000
Q12	,5110	,4051	,5540	,6256	,5264
Q13	,3395	,3472	,6194	,4863	,4264
Q14	,4592	,4131	,4953	,5072	,5193
Q15	,4234	,3691	,5096	,5339	,4661
Q16	,4294	,6228	,6760	,6674	,6548

Q12 Q13 Q14 Q15 Q16

Q12	1,0000				
Q13	,4982	1,0000			
Q14	,5426	,4924	1,0000		
Q15	,6515	,5341	,5175	1,0000	
Q16	,6013	,5550	,5742	,5866	1,0000