

Universidade de Brasília

Instituto de Psicologia

Curso de Pós-Graduação em Psicologia

Avaliação da prática docente no ensino superior:

fundamentação e aplicação

Luiz Francisco Bastos Peres dos Santos

Brasília, DF

2007

Universidade de Brasília  
Instituto de Psicologia  
Curso de Pós-Graduação em Psicologia

Avaliação da prática docente no ensino superior: fundamentação e aplicação

Luiz Francisco Bastos Peres dos Santos

Brasília, DF

2007

Universidade de Brasília  
Instituto de Psicologia  
Curso de Pós-Graduação em Psicologia

Avaliação da prática docente no ensino superior: fundamentação e aplicação

Luiz Francisco Bastos Peres dos Santos

Tese apresentada ao Programa de Pós-graduação em Psicologia do Instituto de Psicologia da Universidade de Brasília como requisito parcial à obtenção do grau de Doutor em Psicologia.

Orientador: Jacob A. Laros

Brasília, DF

Mai de 2007

Avaliação da prática docente no ensino superior: fundamentação e aplicação

Tese aprovada pela banca examinadora constituída por:

---

Prof. Jacob A. Laros, Ph. D. (Presidente)  
Instituto de Psicologia/Universidade de Brasília

---

Prof. Denise de Souza Fleith, Ph.D. (Membro)  
Instituto de Psicologia/Universidade de Brasília

---

Prof. Dr. Henrique Flávio Rodrigues da Silveira (Membro)  
Centro Universitário do Distrito Federal/Banco Central do Brasil

---

Prof. Luiz Pasquali, Docteur (Membro)  
Instituto de Psicologia/Universidade de Brasília

---

Prof. Dr. Mauro Luiz Rabelo (Membro)  
Departamento de Matemática/Universidade de Brasília

---

Prof. Cláudio Vaz Torres, Ph. D. (Suplente)  
Instituto de Psicologia/Universidade de Brasília

## Dedicatória

Aos meus pais Francisco e Aldira (em memória), os primeiros e fundamentais professores em minha vida.

À minha esposa Valéria e ao meu filho Gustavo, professores que me mostram, a cada dia, que a vida é feita de pessoas, relacionamentos e emoções.

Ao meu irmão Rômulo (em memória), por me ensinar o valor de uma amizade que transcende os laços familiares e que, espero, se estenda além dos limites da vida e da morte.

## Agradecimentos

A Deus por ter me permitido chegar até aqui.

Ao Professor Dr. Jacob A. Laros, por sua inesgotável disponibilidade, segura orientação e também pela fraternal convivência que transcendeu a relação orientador-orientando.

Aos membros da Banca Examinadora pela sua gentileza em aceitar o convite para participar da avaliação desta tese.

Aos meus professores do Programa de Doutorado e, em especial aos Professores Doutores Hartmut Günther (UnB) e Luiz Pasquali (UnB), pelos conhecimentos transmitidos e pelos sábios ensinamentos de vida.

À Professora Doutora Denise de Souza Fleith (UnB), por ter me ensinado que eventuais barreiras são também fontes de entusiasmo e imaginação para o pesquisador.

Aos membros da Banca Examinadora do Exame de Qualificação, Professoras Doutoras Fatima Ali Abdalah Abdel Cader Nascimento (UnDF) e Denise de Souza Fleith (UnB), pelas proveitosas sugestões que ajudaram a enriquecer esta tese.

Ao Professor Linaldo José Malveira Alves (Vice-Reitor UnDF) e às Professoras Gizelha Pereira Gomes (UnDF), Simone Cerveira de Castro (UnDF) e Valéria Fachine (UnDF), pela amizade, incentivo permanente e sem cuja ajuda esta tese não poderia ter sido realizada.

Aos colegas Girlene Ribeiro de Jesus, Josemberg Moura de Andrade, Cristiane Faiad e Fábio Iglesias, pelo agradável convívio e por me mostrarem que é sempre possível melhorar a qualidade de nosso trabalho mediante um fraterno e sólido questionamento.

Ao UnDF, nas pessoas do Professor Rezende Ribeiro de Rezende (Magnífico Reitor), Professor Marcos Ribeiro de Azevedo (Diretor da Faculdade de Ciências Gerenciais e de Tecnologias), Maria Consuelene Marques (Diretora da Biblioteca), Maurinete dos Santos (Biblioteca) e Yih Min Huang (CPD), pelo apoio durante a realização desta tese.

À professora Doutora Ana Maria Nogales de Vasconcelos (UnB) e ao Professor Vasco Pedro Moretto, por terem acreditado no meu potencial e me recomendado para ingressar no programa de Doutorado.

Por último, mas não menos importante, aos meus alunos, por me permitirem não só aprender a ensinar, mas sobretudo aprender a aprender.

## SUMÁRIO

Lista de Tabelas .....	7
Lista de Figuras .....	8
Resumo .....	9
Abstract .....	10
Capítulo 1 – Considerações Iniciais .....	11
Capítulo 2 – Revisão da Literatura .....	17
Capítulo 3 – Metodologia de construção de instrumentos de mensuração de construtos ...	28
Capítulo 4 – Método .....	32
4.1 Locus da Pesquisa .....	32
4.2 O Instrumento .....	32
4.3 Procedimentos .....	36
4.4 Participantes .....	37
4.5 Análise dos Dados .....	37
Capítulo 5 – Resultados e Discussão .....	53
5.1 Análise Exploratória .....	53
5.2 Análise Preliminar .....	56
5.3 Análise Confirmatória .....	68
Capítulo 6 – Considerações Finais .....	80
Referências .....	87
Anexo 1: Características dos membros do comitê revisor .....	106
Anexo 2: Modelo do questionário aplicado .....	108
Anexo 3: Modelos dos e-mails enviados aos alunos e professores da IES .....	111
Anexo 4: Orientação para a aplicação dos questionários .....	114

## Lista de Tabelas

Tabela 1. Setor regulamentado da educação formal no Brasil.....	12
Tabela 2. Itens do questionário.....	35
Tabela 3. Respostas dos questionários por curso, sexo e turno.....	37
Tabela 4. Ocorrência de múltipla marcação de respostas para os itens do questionário.....	53
Tabela 5. Ocorrência de valores ausentes para os itens do questionário.....	54
Tabela 6. Assimetria dos itens do questionário.....	55
Tabela 7. Média e desvio-padrão dos itens do questionário.....	56
Tabela 8. Correlações entre os itens do questionário.....	57
Tabela 9. Índice KMO e teste de esfericidade de Bartlett.....	57
Tabela 10. Autovalores empíricos e aleatórios dos primeiros 16 componentes.....	58
Tabela 11. Fatores e percentual de variância explicada.....	60
Tabela 12. Cargas fatoriais e comunalidade dos itens do questionário.....	62
Tabela 13. Cargas fatoriais, comunalidade e correlação item-total para os itens do fator de segunda ordem.....	64
Tabela 14. Cargas fatoriais dos itens que fazem parte do fator 1.....	65
Tabela 15. Cargas fatoriais dos itens que fazem parte do fator 2.....	66
Tabela 16. Média, desvio-padrão e correlação item-total para os itens do fator 1.....	66
Tabela 17. Média, desvio-padrão e correlação item-total para os itens do fator 2.....	67
Tabela 18. Análise confirmatória do modelo de mensuração relativo ao relacionamento professor-aluno.....	72
Tabela 19. Análise confirmatória do modelo de mensuração relativo à prática de ensino.....	75
Tabela 20. Análise confirmatória do modelo estrutural de avaliação do professor pelo aluno.....	78

## Lista de Figuras

Figura 1. Construção e validação de um modelo. ....	21
Figura 2. Modelo teórico de avaliação da prática docente no ensino superior. ....	26
Figura 3. Modelo para a construção de um instrumento. ....	28
Figura 4. Validação semântica dos itens do questionário. ....	34
Figura 5. Limpeza da base de dados. ....	39
Figura 6. Modelo estrutural de avaliação do professor pelo aluno (modelo teórico). ....	69
Figura 7. Modelo de mensuração relativo ao fator relacionamento professor-aluno (modelo teórico). ....	71
Figura 8. Modelo de mensuração relativo ao fator relacionamento professor-aluno. ....	72
Figura 9. Modelo de mensuração relativo ao fator prática de ensino (modelo teórico). ....	73
Figura 10. Modelo de mensuração relativo ao fator prática de ensino. ....	74
Figura 11. Modelo estrutural de avaliação do professor pelo aluno (modelo teórico). ....	76
Figura 12. Modelo estrutural de avaliação do professor pelo aluno. ....	77

## Resumo

Várias décadas de pesquisa sobre a efetividade do ensino mostram que a avaliação do professor pelo aluno, bem planejada, construída e administrada, pode contribuir de forma significativa para a melhoria da qualidade do ensino. Essa modalidade de avaliação é uma das mais efetivas para se alcançar um aprimoramento dos processos de ensino e aprendizagem. Nesta tese foi desenvolvido um modelo multifatorial, representativo da avaliação pelo aluno, da prática docente do professor de ensino superior. Um instrumento foi desenvolvido para validar empiricamente o modelo. Após três estudos pilotos, o instrumento foi aplicado em uma Instituição de Educação Superior do Distrito Federal, envolvendo alunos de 654 turmas e 227 professores. A análise exploratória do banco de dados, levando-se em conta os casos de múltipla marcação de respostas, valores ausentes e valores extremos, resultou numa base de trabalho com 14.951 observações. Análises fatorial exploratória e confirmatória revelaram a existência de dois fatores: prática de ensino (9 itens,  $\alpha = 0,96$ ) e relacionamento professor-aluno (7 itens,  $\alpha = 0,96$ ). Com base nos resultados obtidos, são oferecidos subsídios para que os professores de ensino superior possam refletir sobre a melhoria da qualidade de sua prática docente.

Palavras-chave: Análise fatorial; avaliação do professor pelo aluno; construção de instrumento; equações estruturais; qualidade do ensino.

## Abstract

Various decades of research on the effectiveness of teaching indicate that evaluation of teachers by their students, provided that this occurs in a well planned, constructed and administered way, can contribute significantly to the improvement of the quality of teaching practices. This kind of evaluation is one of the most effective ways to achieve an increase in the quality of learning and teaching process. The main purpose of the present study is the development and validation of a multifactorial model indicating which main factors affect the instructional quality of teachers at institutions of higher education. To validate this model an instrument was developed. After three pilot studies, the instrument was administered to students of 654 classrooms and 227 teachers in a private Brazilian institution of higher education. An exploratory analysis of multiple answers, missing values and outliers, resulted in a database with 14,951 observations. The exploratory and confirmatory factorial analysis showed two factors: teaching practice (9 itens,  $\alpha = 0.96$ ) and teacher-student relation (7 itens,  $\alpha = 0.96$ ). Finally, based on the obtained results, suggestions are offered to improve the quality of teaching practices of teachers at institutions of higher education.

Key-words: factor analysis; evaluation of teachers by students; construction of an instrument; structural equations; quality of teaching.

## Capítulo 1 – Considerações Iniciais

A educação pode ser considerada como uma prática fundamental da espécie humana, distinguindo o modo de ser cultural dos homens do modo natural de existir dos demais seres vivo. Tal prática, com profunda influência na própria existência do homem, nos leva a considerar a escola tanto como local e ponto de partida, como também universal e ponto de chegada (Gadotti, 2003). A universalidade do pensar sobre a educação pode ser apreciada, em toda a sua plenitude, debruçando-se sobre o estudo das idéias pedagógicas ao longo do tempo. Este estudo mostra que as diversas correntes de pensamento (oriental, grego, romano, medieval, renascentista, moderno, iluminista, positivista, socialista, etc) constituem um amplo espectro de visão dos pensadores que discutiram as questões educacionais e revela a forma dialética em que se desenvolveu o movimento do pensamento pedagógico até a práxis atual.

Nesta práxis não se pode deixar de considerar as rápidas e radicais transformações, no final do século passado e início do novo milênio, que afetaram e continuam afetando profundamente o mundo e a sociedade. Destas transformações são mais visíveis o aumento do conhecimento tecnológico e a transformação do processo produtivo.

Estas transformações acarretam conseqüências sociais imediatas tais como a extinção de milhares de empregos, o aumento da economia informal, o trabalho infantil, a evasão escolar e a desestruturação familiar. Neste contexto de agravamento dos problemas sociais é preciso pensar estrategicamente. Porém, pensar estrategicamente é, antes de mais nada, pensar em educação (Trigueiro, 2000). A educação, apesar de ser uma das maiores consumidoras de dinheiro público e empregadora de uma grande força de trabalho, revela uma unanimidade: é difícil encontrar alguém, dentro ou fora do sistema de educação formal, que esteja satisfeito com o seu desempenho (Egan,2002).

O sistema educacional brasileiro é constituído pelos setores de cursos livres (idiomas, cursinhos pré-vestibulares, preparatórios para concursos, cursos técnicos, treinamento corporativo, etc) e pelo setor regulamentado (educação formal), conforme ilustrado na Tabela 1.

Tabela 1. Setor regulamentado da educação formal no Brasil.

nível de ensino	tipos/modalidade	órgão regulador
Creche/infantil	público/privado	Secretaria Municipal
Ensino fundamental	público/privado/rede	Secretaria Estadual
Ensino médio	público/privado/rede	Secretaria Estadual
Ensino superior	público /privado /rede/ sequencial/tecnólogo/ on-line	Ministério da Educação e Cultura (MEC)
Pós-graduação <i>stricto sensu</i>	público /privado /on-line	Ministério da Educação e Cultura (MEC)

Fonte: Souza (2005).

A Tabela 1 nos mostra que o sistema de ensino formal brasileiro possui um caráter extremamente descentralizado e é constituído pela interligação complexa de vários subsistemas de gestão e execução. Tal constituição representa um grande desafio para todos os envolvidos na área educacional, dadas as características de nosso país:

- O Brasil, com mais de 8,5 milhões de quilômetros quadrados de extensão territorial, tamanho equivalente a quase toda a Europa, excetuando-se a porção russa, é o quarto país do mundo em terras contínuas, sendo suplantado apenas pela Rússia, China e Canadá (Souza, 2005).
- Em relação à população, o Brasil é o quinto país mais populoso do planeta, ficando atrás apenas da China, Índia, Estados Unidos e Indonésia (Souza, 2005).
- O processo de ocupação do território brasileiro, por diversos grupos culturais e étnicos, e as distintas atividades econômicas que os acompanharam, acarretaram para as regiões brasileiras uma diferenciação histórica e heterogeneidade cultural, à qual se acrescentou uma grande desigualdade econômica e social (Souza, 2005). As origens desta desigualdade são atribuídas, em grande parte, ao atraso educacional de décadas e à baixa escolaridade média da população brasileira (MEC, 2001; Souza, 2005).

Toda essa heterogeneidade econômica, social e cultural do Brasil encontra-se agrupada em um regime federativo e republicano, constituído pelo Distrito Federal e pelos estados, além dos municípios em que os estados são subdivididos. A cada uma dessas unidades federativas são atribuídas, pela Constituição Federal, responsabilidades diversas em todos os planos, inclusive o educacional. Cabe ao governo federal, em relação à educação básica,

um papel normativo, redistributivo e supletivo, atuando com auxílios e subsídios para diminuir desigualdades sociais e regionais (Souza, 2005).

Este papel sugere um cenário preocupante pois em nosso país, muitas vezes, a regra geral é calcada numa aliança entre a interferência política e um sistema complexo e não transparente de recursos, além da primazia desta aliança sobre as razões de ordem técnica (Souza, 2005). Essa preocupação é agravada pelo fato de que todo o sistema de ensino formal no Brasil, exceto o ensino superior, é majoritariamente público: no ano 2000, 88% dos alunos matriculados na educação básica eram atendidos por escolas públicas, enquanto no ensino superior este percentual era de apenas 35% (MEC, 2001).

Do ponto de vista legal cabe ao governo federal avaliar a qualidade do ensino em todo o país e em todos os níveis. Para atender a este requisito foram concebidos os diversos sistemas oficiais de avaliação educacional no Brasil: Sistema Nacional de Avaliação da Educação Básica, Exame Nacional do Ensino Médio e Exame Nacional de Cursos Superiores (Souza, 2005).

O Sistema Nacional de Avaliação da Educação Básica (SAEB) foi criado e aplicado em caráter experimental no início da década de 1990, contando com o apoio financeiro do Banco Mundial, e tem por objetivo avaliar a qualidade da educação básica oferecida pelos sistemas de ensino existentes no Brasil, oferecendo um diagnóstico confiável da qualidade da educação brasileira, estado por estado (Andrade, 2005; Jesus, 2004; Souza, 2005). A partir de 1995 o SAEB passou a ser regularmente realizado a cada dois anos. Ele consiste na utilização de questionários dirigidos a professores e diretores, além de testes de rendimento aplicados a uma amostra de alunos das 4<sup>as</sup> e 8<sup>as</sup> séries do ensino fundamental e da 3<sup>a</sup> série do ensino médio (Souza, 2005).

Os resultados do SAEB procuram identificar o que os alunos sabem e são capazes de fazer nos diferentes momentos do seu percurso escolar e o que seria desejável que eles soubessem (Souza, 2005). Esses dados propiciam ao MEC e às secretarias estaduais e municipais de educação, a definição de ações voltadas para a construção de uma escola de qualidade. O SAEB foi complementado, em algumas unidades da federação, por sistemas locais, entre os quais podem ser citados o Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (SARESP) e o Programa de Avaliação da Rede Pública de Educação Básica (PROEB), em Minas Gerais.

Em 2005 o MEC criou a Avaliação Nacional da Educação Básica (ANEB), que é similar ao antigo SAEB, e a Avaliação Nacional do Rendimento Escolar (ANRESC), mais

extensa e detalhada. A ANRESC destina-se a levantar informações sobre o desempenho, em leitura e matemática, de cada uma das 43 mil escolas urbanas de 4<sup>as</sup> e 8<sup>as</sup> séries da rede pública brasileira, em mais de 5 mil municípios do País.

O Exame Nacional do Ensino Médio (ENEM) destina-se aos alunos que estão concluindo ou que já concluíram o ensino médio e tem como ênfase a avaliação do perfil de saída dos egressos deste nível de ensino. Seu objetivo principal é a avaliação do desempenho dos alunos, ao término da escolaridade básica, segundo uma estrutura de competências associadas aos conteúdos das diversas disciplinas, que se espera tenha sido incorporada pelo aluno para fazer frente aos constantes desafios da vida moderna (Souza, 2005).

O ENEM oferece uma medida dos resultados que a escola apresenta diante dos desafios impostos pelos mecanismos estruturais da sociedade, tanto no tocante ao pleno exercício da cidadania, como no sentido de prover uma formação adequada para o nível superior de ensino (Souza, 2005). Os resultados do ENEM já são usados por algumas Instituições de Educação Superior (IES) do país como complementares aos tradicionais processos vestibulares ou mesmo como substitutos destes processos, sendo também utilizado como um dos critérios de acesso ao Programa Universidade para Todos (PROUNI) do MEC.

O Exame Nacional de Cursos Superiores (ENC), que foi tornado obrigatório a partir de 1996 para todos os alunos formandos das IES públicas e privadas, foi chamado pelo alunado de Provão. Esta denominação foi incorporada por toda a sociedade brasileira, incluindo os órgãos oficiais de educação, que a adotaram, inclusive, como título de uma revista de divulgação dos seus pressupostos e objetivos (Vianna, 2003).

O Provão nasceu sob o signo da contestação de alguns segmentos, inclusive alunos e professores, que propuseram um boicote ao exame, o qual, no entanto, foi declinando ao longo dos anos, passando de 11,3% em 1996 para 1,8% em 2002 (Souza, 2005). O número de cursos avaliados passou de 616 em 1996 para 5.030 em 2002 e o número de inscritos pulou de 56 mil em 1996 para 396 mil em 2002, representando cerca de 90% do total de alunos graduados em todas as carreiras no Brasil nesse ano (Souza, 2005).

O amplo processo de divulgação dos resultados (divulgação na imprensa; disponibilização dos dados no site do MEC; preparação de informações individualizadas por cursos, enviadas aos coordenadores; realização de seminários anuais, por área, com a presença dos coordenadores de curso), contribuiu para que o Provão se tornasse um

processo consolidado (Souza, 2005). Este processo foi aceito pela sociedade brasileira que, aliando-se ao governo, passou a exigir das IES a demonstração de um diferencial de qualidade (Souza, 2005). Esta exigência resultou, em muitos casos, na contratação de bons professores, desenvolvimento de melhores programas acadêmicos e investimento em boas instalações físicas, laboratórios e bibliotecas, melhorando, de forma substancial, a qualidade do ensino superior (Souza, 2005).

O Exame Nacional do Estudante (ENADE) foi instituído pelo MEC em 2004, como substituto do Provão. O ENADE ainda não possui dados suficientes para se analisar os aspectos positivos ou negativos de sua implementação em substituição a uma iniciativa vitoriosa e amplamente aceita pela sociedade brasileira que era o Provão.

O complexo sistema de avaliação em todos os níveis de ensino, implantado no Brasil em pouco mais de uma década, continua enfrentando desafios. Entre tais desafios podem ser citados: o estabelecimento de padrões de desempenho escolar, passíveis de comparação num quadro de grande diversidade regional; aperfeiçoamento dos sistemas estaduais de avaliação; aumento da precisão na identificação e ponderação dos fatores, internos e externos à escola, que impactam o desempenho dos alunos. A incidência direta dos fatores econômicos, sociais e culturais, sobre o desempenho dos alunos da educação básica, já foi demonstrado em alguns estudos recentes (Andrade, 2005; Fukuda, 2003; Jesus, 2004; Jesus & Laros, 2004). Existe, portanto, uma grande margem de ação no ensino superior, que é propriamente educativa, e que se desenvolve em sala de aula.

Sendo o corpo docente de uma IES a principal fonte do conhecimento acadêmico, podemos considerar a qualidade do ensino como determinada, basicamente, pelas ações que norteiam os processos de ensino e aprendizagem, e os alunos como os principais fornecedores de informação sobre a efetividade do professor (Bem, 2004). Assim, a avaliação sistemática pelo aluno da prática docente passa a ter importância primordial para o aprimoramento desta prática e para a melhoria da qualidade do ensino ministrado nas IES.

É neste contexto, de utilização dos resultados obtidos com a avaliação para o aperfeiçoamento da prática docente e para a elevação da qualidade do ensino nas IES, que se insere o presente trabalho. Para tanto, fazemos um recorte na complexidade do estudo do ensino superior no Brasil, apresentando uma proposta de natureza metodológica, para a avaliação da qualidade deste ensino, no contexto da sala de aula.

Esta proposta, além de contribuir para o incremento da literatura nacional sobre o tema e envolver os dois principais atores dos processos de ensino e aprendizagem – o professor e o aluno-, tem os seguintes objetivos:

#### Geral

- Desenvolver e validar um modelo de avaliação, pelo aluno, da prática docente no ensino superior.

#### Específicos

- Construir um instrumento para validar empiricamente o modelo desenvolvido.
- Analisar, com base em técnicas estatísticas multivariadas, as propriedades psicométricas do instrumento construído.
- Oferecer subsídios para que os professores de ensino superior possam refletir sobre a melhoria da qualidade de sua prática docente.

## Capítulo 2 – Revisão da Literatura

Avaliação da qualidade do ensino é um tema que, além de representar uma preocupação quase que universalmente presente em praticamente todo debate sobre educação, revela uma unanimidade: sua operacionalização é considerada uma tarefa de grande complexidade (Pasquali, 1984). Apesar disto, a pesquisa científica sobre a temática da avaliação do ensino pelo aluno nos forneceu uma rica e volumosa literatura internacional, predominantemente de origem estadunidense, com mais de 1.500 títulos (Cashin, 1995).

Alguns dos resultados oferecidos nos estudos revisados sobre a avaliação da prática docente são:

- Os alunos atribuem uma avaliação mais alta para os professores que para as professoras (Atamian & Ganguli, 1993; Basow, 1994; Basow & Silberg, 1987; Ferber & Huber, 1975; Goldberg & Callahan, 1991; Haemmerlie & Highfill, 1991; Sidanius & Crane, 1989).
- As professoras recebem avaliações melhores que os professores (Kierstead, D'Agostino & Dill, 1988; Tatro, 1995).
- O professor que demonstra um conjunto de características femininas e masculinas recebe dos alunos uma avaliação maior que a que receberia se exibisse características de apenas um dos gêneros (Freeman, 1994; Martin, citado por Andersen & Miller, 1997).
- A grande maioria dos estudos conclui que não há diferença nas avaliações dos professores feitas por alunos ou alunas (Basow & Distenfeld, 1985; Bennet, 1982; Centra & Gaubatz, 2000; Doyle & Whitely, 1974; Dukes & Victoria, 1989; Elmore & LaPointe, 1974; Feldman, 1993; Freeman, 1994; Goodwin & Stevens, 1993; Harris, 1975; Isaacson & cols., 1964; Ludwig & Meacham, 1997).
- Tem-se uma correlação positiva entre a avaliação do professor pelo aluno e as notas obtidas por eles (Goldberg & Callahan, 1991; Nimmer & Stone, 1991; Sailor, Whorten & Shin, 1997; Scherr & Scherr, 1990; Wilson, 1998).
- A correlação entre a avaliação do professor pelo aluno e as notas obtidas por eles é praticamente nula (25 estudos relatados por Aleamoni & Hexner, citado por

Aleamoni, 1999; Baird, 1987; Centra, 2003; Gigliotti & Buchtel, 1990; Howard & Maxwell, 1980).

- Alunos de pós-graduação e alunos veteranos tendem a avaliar de forma mais favorável os professores que os alunos calouros e novatos (18 estudos relatados em Aleamoni & Hexner, citado por Aleamoni, 1999; Moritsch & Suter, 1988).
- As relações entre o *status* do aluno (calouro, veterano, graduando) e a avaliação atribuída ao professor pelo aluno são não significativas (oito estudos em Aleamoni & Hexner, citado por Aleamoni, 1999).
- A correlação entre o *status* acadêmico do professor (instrutor, assistente, associado) e as avaliações dos alunos não é significativa (Aleamoni & Graham, 1974; Aleamoni & Ymer, citado por Aleamoni, 1999; Heilman & Armentrout, 1936; Linsky & Straus, 1975).
- Os professores de maior *status* acadêmico são mais bem avaliados que os professores de menor *status* (Downie, 1952; Gage, 1961).
- Os professores de classes maiores tendem a receber avaliações piores que os professores de classes menores (oito estudos relacionados em Aleamoni & Hexner, citado por Aleamoni, 1999; Cohen, 1983; Crittenden, Norr & LeBailly, 1975; Mateo & Fernandez, 1996; Watkins, 1990).
- Não foi encontrada uma relação significativa entre o tamanho da classe e as avaliações dos alunos (sete estudos relacionados em Aleamoni & Hexner, citado por Aleamoni, 1999; Anderson & Guichard, 1988; Lin, citado por Aleamoni, 1999; Shapiro, 1990).

A revisão da literatura nos revela, portanto, a existência de resultados conflitantes sobre a importância de variáveis usualmente utilizadas em avaliações, pelo aluno, do trabalho do professor de ensino superior. Entre estas variáveis estão o gênero, nota do aluno, *status* do aluno, *status* acadêmico do professor e tamanho da classe.

A avaliação do professor pelo aluno, ao longo de seu percurso histórico, teve também que superar uma série de objeções, por parte dos professores que se posicionaram contra tal procedimento (Aleamoni, 1999; Hildebrand, 1972):

- Objeção 1: Os alunos, devido à sua imaturidade e inexperiência, não podem fazer julgamentos consistentes sobre o professor.

Réplica: diversos estudos nos indicam exatamente o contrário: correlações entre avaliações dos alunos dos mesmos professores e cursos variam de 0,70 a 0,87,

enquanto que a estabilidade das avaliações dos alunos de um ano para o outro apresentam correlações na faixa de 0,87 a 0,89 (Costin, Greenough & Menges, 1971; Hativa, 1996; Hogan, 1973).

- Objeção 2: Somente colegas com excelente registro de publicações e nível de *expertise* são qualificados para ensinar e avaliar o ensino de seus pares.

Réplica: este posicionamento encontra eco no entendimento que apregoa que bom ensino e boa pesquisa estão tão intimamente relacionados que é desnecessário avaliá-los independentemente. Encontramos, no entanto, estudos que revelam fraca correlação positiva entre produtividade de pesquisa e efetividade de ensino (Maslow & Zimmerman, 1956; McGrath, 1962) e estudos que mostram uma relação não significativa entre produtividade de pesquisa do professor e avaliação do aluno (Guthrie, 1949; Hayes, 1971; Linsky & Straus, 1975; Voeks, 1962).

- Objeção 3: As avaliações de alunos nada mais são que concursos de popularidade, em que os professores mais descontraídos, amistosos e mais engraçados são sempre os vencedores.

Réplica: a literatura nos revela exatamente o contrário. Os alunos são juízes discriminatórios do ensino efetivo, uma vez que professores descontraídos, amistosos e engraçados, recebem avaliações baixas se seu ensino não for bem organizado e seus métodos conduzirem a baixo nível de aprendizagem dos alunos (Marlin, 1987; Tollefson, Chen & Kleinsasser, 1989; Ware & Williams, 1977; Waters, Kemp & Pucci, 1988).

- Objeção 4: Os alunos somente serão capazes de avaliar o professor após muitos anos de formados.

Réplica: os estudos mostram que são semelhantes as avaliações de professores por alunos que estão cursando a graduação e por alunos já graduados (Aleamoni & Yimer, citado por Aleamoni, 1999; Druckmer & Remmers, 1951; Marsh, 1977; Marsh & Overall, 1979).

- Objeção 5: Os instrumentos utilizados na avaliação dos professores pelos alunos são inválidos e não confiáveis.

Réplica: esta afirmação não se sustenta, uma vez que os estudos revelam que instrumentos bem planejados, construídos e aplicados, possuem uma alta consistência interna, confiabilidade e validade (Abbot & Perkins, 1978; Abrami, d'Apollonia & Cohen, 1990; Baird, 1987; Burdsal & Bardo, 1986; Costin,

Greenough & Menges, 1971; Ellet, Loup & Culross, 1997; Gigliotti & Buchtel, 1990; Harrison, Ryan & Moore, 1996; Koon & Murray, 1995; Marsh, 1984; Nimmer & Stone, 1991; Pasquali, 1984; Prave & Baril, 1993; Prosser & Trigwell, 1990).

- Objeção 6: As avaliações dos alunos não podem ser usadas para a melhoria do ensino.

Réplica: os estudos mostram uma melhora no ensino dos professores que receberam uma consultoria particular com base no *feedback* fornecido pelas avaliações dos alunos (McKeachie, 1979; Schmelkin & Spencer, 1997; Stevens & Aleamoni, citado por Aleamoni, 1999).

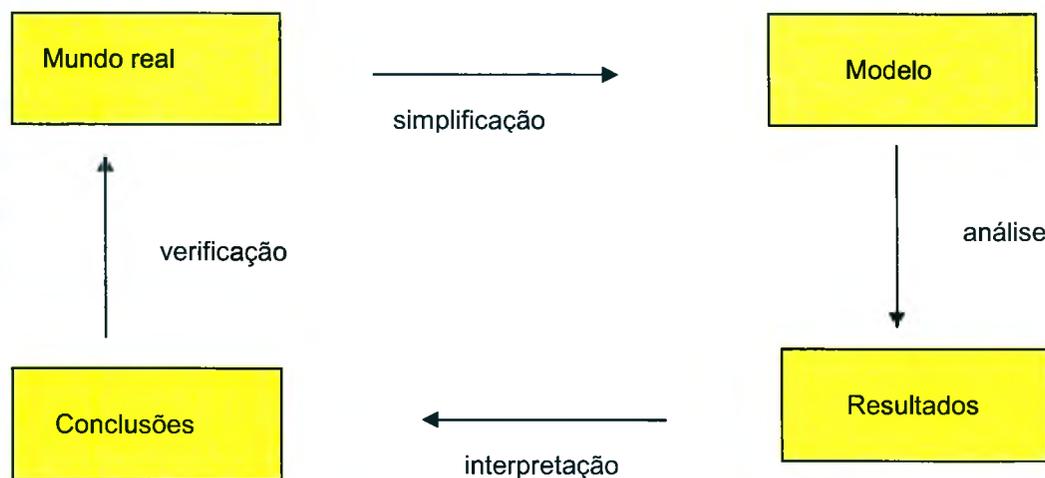
- Objeção 7: As notas que os alunos recebem no curso são altamente correlacionadas com a avaliação que eles fazem dos professores.

Réplica: os estudos de natureza correlacional não são conclusivos quanto a esta afirmação. Assim é que alguns estudos relataram a correlação entre as variáveis nota do aluno e avaliação do professor pelo aluno como nula (Aleamoni & Hexner, citado por Aleamoni, 1999; Baird, 1987; Gigliotti & Buchtel, 1990), outros relataram esta correlação como positiva (Aleamoni & Hexner, citado por Aleamoni, 1999; Goldberg & Callahan, 1991; Nimmer & Stone, 1991; Sailor, Worthen & Shin, 1997; Scherr & Scherr, 1990), enquanto outro estudo relatou como negativa essa correlação (Sailor, Worthen & Shin, 1997).

Apesar dessas objeções, representadas, na maioria das vezes, por simples opiniões ou arraigadas posições, sem nenhum embasamento nas pesquisas realizadas, a análise da literatura nos permite considerar a avaliação do professor pelo aluno, quando bem planejada, construída e administrada, como um elemento importante para a melhoria da qualidade do ensino. Além disso, essa revisão nos permite concluir, com relativa segurança, que os alunos, considerados os melhores juízes para avaliação do ensino que recebem, possuem capacidade de realizar a avaliação (Scriven, 1995) e, portanto, não devem ser considerados participantes mudos na melhoria da qualidade dos processos de ensino e aprendizagem.

Para que se possa avaliar essa contribuição, torna-se, necessário, inicialmente, o estabelecimento de um quadro de referência que sirva de orientador da discussão e ofereça os parâmetros que facilitem o posicionamento da argumentação. O quadro de referência a ser estabelecido nesta tese será um modelo de avaliação, pelo aluno, da prática docente no

ensino superior. O processo de construção e validação do modelo está ilustrado na Figura 1.



**Figura 1.** Construção e validação de um modelo.

No processo ilustrado na Figura 1, a simplificação é necessária, uma vez que, se um modelo não fosse uma simplificação ele não seria um modelo: seria uma cópia. Na análise do modelo são empregadas de forma adequada as ferramentas de natureza estatística ou matemática colocadas à disposição do pesquisador. Os resultados obtidos são, então, interpretados à luz das bases substantivas que levaram à construção do modelo. Finalmente, as conclusões obtidas são verificadas, levando-se em conta que o modelo não fornece, necessariamente, todas as informações relevantes que o pesquisador necessita. As conclusões são empregadas, sobretudo, visando a validação e adequação do modelo. Finalmente, com base nos resultados obtidos com o processo, mantém-se o modelo ou prossegue-se com o seu refinamento mediante um novo ciclo de pesquisa. O modelo construído nesta tese teve como base os modelos relevantes encontrados na literatura, descritos a seguir.

Aleamoni (1978) descreve o processo de construção e validação fatorial do modelo conhecido como *Course Instructor Evaluation Questionnaire* (CIEQ), desenvolvido na

Universidade do Arizona, em Tucson, após mais de 15 anos de pesquisa e experiência na área de avaliação instrucional. Em 1961, com base na literatura existente na época, foram coletados cerca de 150 itens, relacionados à prática do professor em sala de aula, que, após a contribuição de um comitê conjunto de alunos e professores da Universidade Estadual da Pensylvania, foi ampliado para 1.000 itens.

Este *pool* foi refinado após sucessivas aplicações a amostras distintas de alunos e cursos e aplicações de análises fatoriais, até resultar, em 1965, num instrumento com 50 itens, ancorados numa escala tipo Likert de 5-pontos (concordo totalmente, concordo, em dúvida, discordo, discordo totalmente). Este questionário forneceu cinco fatores, rotulados de atitude geral em relação ao curso, método de instrução, conteúdo do curso, interesse e atenção, instrutor e itens específicos.

Este modelo permaneceu em uso no período de 1965 a 1972, até que uma outra análise, obtida de 5.346 cursos, gerou um novo questionário com 21 itens e com tempo aproximado de aplicação de 10 minutos. De 1972 a 1978, o questionário foi aplicado a cerca de 5.537 cursos e 470.000 alunos e teve sua consistência interna, variando de 0,88 a 0,98 nos cinco fatores e de 0,81 a 0,94 nos itens.

Pasquali (1984), no desenvolvimento do seu Questionário de Avaliação da Docência (QAD), aplicou um questionário com 92 itens a 354 estudantes universitários do Distrito Federal, sendo 115 do gênero masculino, 162 do gênero feminino e 77 que não se identificaram quanto ao gênero. Foi utilizada uma escala tipo Likert de 7-pontos, que indicava se a proposição expressava algo de característico sobre o professor e/ou a situação de aprendizagem, indo de extremamente característico (grau 7) a nada característico (grau 1). As instruções do questionário pediam aos sujeitos para responderem às proposições com relação a um professor e/ou situação de aprendizagem ideais.

A análise dos dados foi feita por meio da técnica de análise fatorial com componentes principais e rotação oblíqua. Após as análises iniciais, o número de itens foi reduzido de 92 para 71, agrupados em sete fatores, rotulados de o grande educador, desrespeito, laboratório, bom docente, compreensão da disciplina por parte do aluno, respeito pelo conteúdo da disciplina e imparcialidade.

O primeiro fator, o grande educador, é constituído por 31 itens e representa, essencialmente, o perfil de um professor cuja preocupação central consiste em educar alunos que, eventualmente, sejam capazes de se movimentar livre e sabiamente no campo do saber e da própria vida. O fator desrespeito é constituído por 10 itens e representa o

professor negligente e displicente que agride seus alunos. O fator laboratório é constituído por 8 itens que caracterizam a situação de aprendizagem em sua parte prática. O fator bom docente é constituído por 21 itens que caracterizam um professor entusiasmado pela disciplina e capaz de transmitir o saber com clareza, autoridade e respeito. O fator compreensão da disciplina por parte do aluno é constituído por 28 itens. O fator respeito pelo conteúdo da disciplina é constituído por 15 itens que caracterizam um professor capaz de aceitar críticas e reconhecer erros, claro e preciso nas tarefas propostas, preocupado com a participação de todos os alunos e que manifesta respeito pelo aluno como pessoa. O fator imparcialidade é constituído por 21 itens que caracterizam um professor que, embora tenha seus pontos de vista pessoais, é capaz de discutir com convicção, relevância e clareza sobre pontos de vistas contrários aos seus, deixando, inclusive, aos alunos, a decisão sobre a solução mais adequada dos problemas. Estes sete fatores foram, posteriormente, agrupados em três fatores de segunda ordem, rotulados de educação e aprendizado, relacionamento professor-aluno e laboratório.

Moreira (1986), em sua tese de doutorado junto à Faculdade de Educação da Universidade de São Paulo (USP), efetuou uma pesquisa realizada no 2º semestre de 1982, com 1.365 alunos de 72 professores das Faculdades Associadas de São Paulo (FASP), em 106 disciplinas. Foi utilizado um questionário com 13 itens, numa escala tipo Likert de 2-pontos, representando a existência ou não da condição expressa na proposição. Por meio de uma análise fatorial, o autor obteve cinco fatores, aos quais denominou de Fator 1, Fator 2, ..., Fator 5. O mesmo questionário foi aplicado em 1985 na mesma instituição a 81 professores e 116 disciplinas, e resultou em 3 fatores, que foram denominados de Fator 1, Fator 2 e Fator 3.

Macedo (2001), em sua tese de doutorado junto ao programa de Engenharia da Produção da Universidade Federal de Santa Catarina (UFSC), utilizou um questionário com 16 itens, representando três grandes objetivos da ação docente: domínio técnico e científico, habilidade de ensino e desempenho pessoal e situacional. Os itens foram avaliados, numa escala tipo Likert de 5-pontos (sempre, quase sempre, na média, raramente, nunca), por 2.980 alunos da Universidade do Vale do Itajaí (UNIVALI), representando 90% de participação e envolvendo 419 turmas de 175 docentes. Os professores tiveram seu desempenho avaliado, nas três áreas propostas, por meio de técnicas de análise fatorial de correspondência múltipla.

Bem (2004), em sua tese de doutorado junto ao programa de Engenharia da Produção da UFSC, utilizou um questionário com 65 itens representantes, na visão do autor, dos aspectos de qualidade da Unisul (24 itens), gerência do curso de graduação (7 itens), desempenho do docente (31 itens) e autopercepção dos alunos (3 itens). O desempenho docente foi subdividido em oito aspectos: exigência, avaliação, conteúdo, método, relacionamento professor-aluno, produtividade, ética e relação professor-curso. Os itens foram avaliados, numa escala tipo Likert de 4-pontos (sim, parcialmente sim, parcialmente não e não), no 2º semestre de 2002, por 10.802 alunos, englobando 2.511 disciplinas. Utilizando análise fatorial, o autor encontrou quatro fatores para o desempenho do professor, aos quais rotulou de competência, relacionamento e ética, motivação dos discentes e avaliação.

Santos (2004), em sua tese de mestrado junto ao programa de Engenharia da Produção da UFSC, utilizou um questionário com 34 itens representantes, na visão do autor, dos aspectos de organização didático-pedagógica e infra-estrutura do curso, auto-avaliação do aluno e desempenho docente. Os itens foram avaliados, numa escala tipo Likert de 5-pontos (nunca, raramente, na média, quase sempre, sempre), em novembro de 2001, por 1.142 alunos da Faculdade Estadual de Ciências Econômicas de Apucarana (FECEA). Os resultados foram obtidos mediante o emprego de análise fatorial e análise de agrupamentos, para cada um dos conjuntos de questões em cada aspecto. O trabalho de Valério (2004), em sua tese de mestrado junto ao programa de Engenharia da Produção da UFSC, é bastante semelhante ao de Santos (2004), no tocante ao instrumento, local de aplicação e metodologia de análise utilizada, sendo também utilizada a versão do questionário para efeito de auto-avaliação docente.

Os estudos considerados os mais abrangentes sobre a avaliação pelos alunos foram os realizados por Marsh e seus colaboradores, nas décadas de 1980 e 1990 do século passado (Marsh, 1977, 1981, 1982a, 1982b, 1983, 1984, 1986, 1987, 1991a, 1991b, 1993; Marsh & Bailey, 1993; Marsh & Hocevar, 1983, 1984, 1991; Marsh & Overall, 1979; Marsh, Overall & Kesler, 1979a, 1979b; Marsh & Roche, 1997). Nesses estudos foi aplicado o *Student Evaluations of Educational Quality* (SEEQ) a cerca de 50.000 cursos e mais de 1 milhão de estudantes dos Estados Unidos e Austrália.

Esses estudos, além de demonstrarem a confiabilidade e a estabilidade da avaliação, revelaram que esta avaliação é, antes de tudo, uma função do professor e não do curso, além de ser um instrumento útil para aprimorar a qualidade do ensino. O questionário

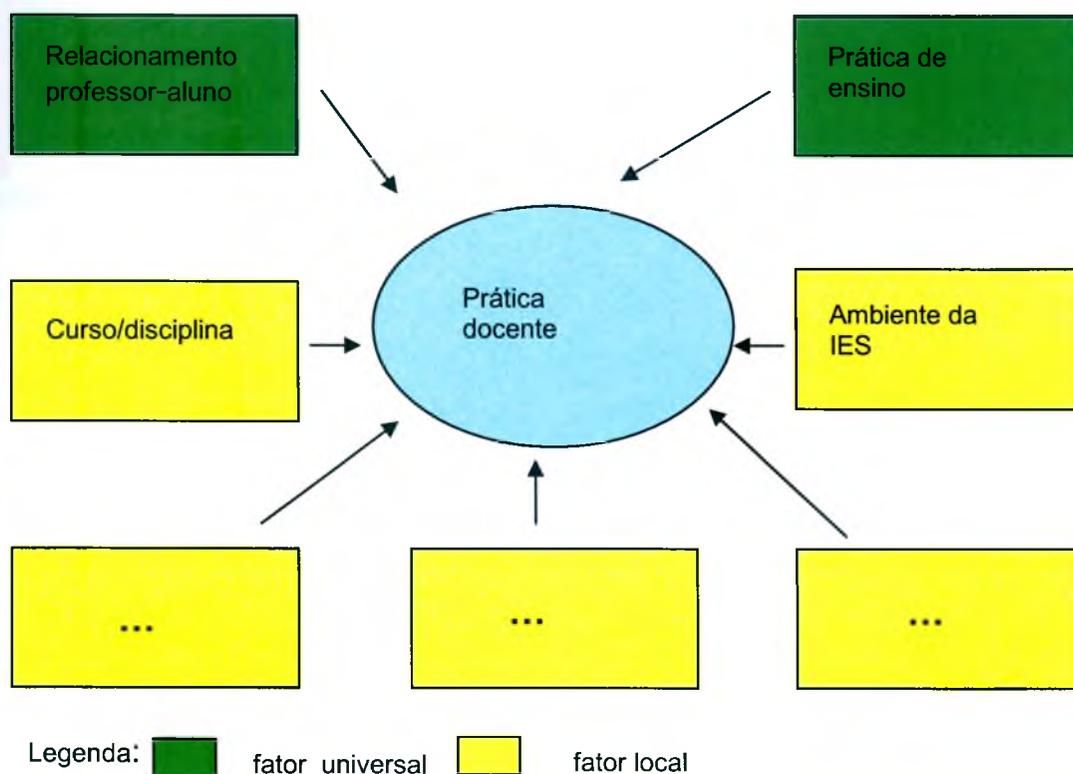
SEEQ, composto por 35 itens, utiliza uma escala tipo Likert de 5-pontos e apresenta uma estrutura fatorial de nove dimensões, rotuladas por Marsh e seus colaboradores de aprendizado/valores, entusiasmo, organização, interação grupal, relacionamento individual, avaliação/notas, tarefas, carga de trabalho/dificuldade e amplitude de cobertura.

A análise da literatura nos revela que a avaliação do professor pelo aluno é usualmente realizada com a utilização de um questionário impresso. A maioria dos questionários são longos e com um grande número de questões. A escala preferida é uma escala do tipo Likert, ancorada num número ímpar de pontos, usualmente 5 ou 7, variando de discordo totalmente a concordo totalmente. Os instrumentos quando bem planejados, construídos e administrados, são válidos e confiáveis.

O construto avaliado, independentemente do rótulo que lhe seja atribuído (efetividade do ensino, desempenho do professor, etc.) é multifatorial. Devido a diferenças de natureza metodológica, não há um consenso quanto ao número de fatores nem quanto à interpretação do significado destes fatores. A maioria dos estudos procura extrair o máximo possível de fatores.

Com base na análise crítica dos modelos encontrados na literatura, podemos, portanto, estabelecer o modelo ilustrado na Figura 2 como um modelo teórico da avaliação, pelo aluno, da prática docente no ensino superior. O modelo é multifatorial, o que está em consonância com a literatura.

Os fatores universais sugeridos representam as características de natureza comum a praticamente todos os modelos descritos na literatura: prática de ensino e relacionamento professor-aluno. Estas características englobam as variáveis de caráter geral, representativas de aspectos relevantes de um ensino de boa qualidade, em sala de aula, e aplicáveis a todos os docentes de uma IES. Os fatores locais englobam as variáveis de caráter particular, representativas de aspectos específicos, peculiares de um determinado contexto (curso/disciplina, modelo de ensino, ambiente da IES, etc.). As características universais e locais representam, de algum modo, respectivamente, o pensar universal e o pensar local sobre a educação, referidos no Capítulo 1 desta tese.



**Figura 2.** Modelo teórico de avaliação da prática docente no ensino superior.

O modelo teórico ilustrado na Figura 2, em uma forma mais simples, será considerado como o ponto de partida para o estabelecimento de um quadro de referência que orientará a discussão e oferecerá os parâmetros que facilitam o posicionamento da argumentação desenvolvida nesta tese. Essa forma mais simples considera apenas as características de natureza universal: relacionamento professor-aluno e prática de ensino.

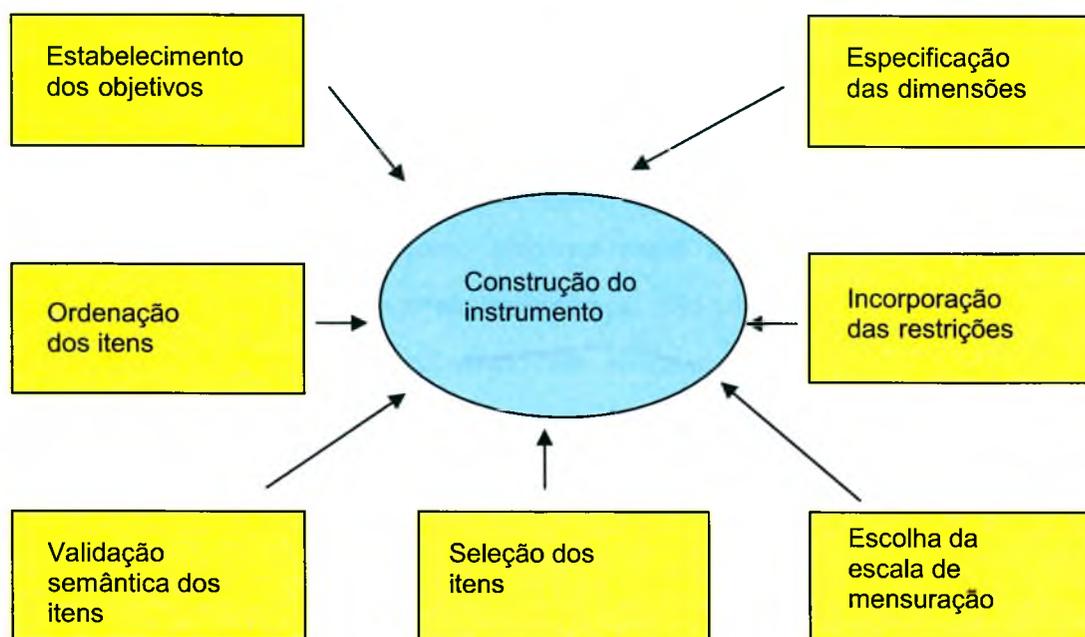
Esta simplificação do modelo inicial foi baseada em duas razões. Em primeiro lugar, consideramos que um modelo, suportado sobretudo por elementos teóricos e evidências empíricas de pesquisas anteriores, deve representar, de forma parcimoniosa, uma estrutura conceitual. Esta afirmação é apoiada, em Psicologia, pelo chamado princípio de economia, que exige que as explicações não sejam mais amplas que o necessário para explicar um evento, devendo ser preferida a explicação mais simples entre as disponíveis (Bachrach,

1969; Pasquali, 2004). Em segundo lugar consideramos imprescindível que um modelo seja passível de ser implementado de forma eficiente e eficaz. Esta implementação deve possibilitar a geração de dados empíricos que corroborem, ainda que de forma provisória, o modelo proposto.

No restante desta tese, ao nos referirmos ao modelo ilustrado na Figura 2, estaremos considerando o modelo na sua versão simplificada, constituído apenas por suas características universais: prática de ensino e relacionamento professor-aluno.

### Capítulo 3 – Metodologia de construção de instrumentos de mensuração de construtos

Para validar o modelo teórico proposto e demonstrar a viabilidade de sua implementação prática, desenvolvemos um instrumento, fundamentado na metodologia ilustrada na Figura 3. Esta metodologia foi construída com base nos trabalhos de Alves-Mazzotti e Gewandszajder (2002), Babbie (2003), Berk (1979), Cozby (2003), Dillman (2000), Finlayson (2002), Fleith e Costa Júnior (2005), Günther (2003), Kaplan (1972), Kerlinger (1980), Malhotra (2004), McGuigan (1976), Oliveira (2001), Pasquali (1999, 2001, 2003, 2004), Posavac e Carey (2004), Wholey, Hatry e Newcomer (2004), Worthen, Sanders e Fitzpatrick (2004) e Wotruba e Wright (1975).



**Figura 3.** Modelo para a construção de um instrumento.

Como o instrumento deve estar em perfeita consonância com o propósito e o foco da avaliação, o claro estabelecimento dos objetivos a serem atingidos deve ser o primeiro passo a ser dado. Desta forma, delinearíamos o que será feito, quem será atingido, que ações serão realizadas com os resultados da avaliação e forneceremos a orientação para a

estrutura e o conteúdo do instrumento. Tal procedimento, além de evitar desperdícios, possibilitará a obtenção de dados úteis. Assim, num instrumento destinado a avaliar a prática docente, uma pergunta sobre “a satisfação do aluno quanto ao preço da alimentação servida na lanchonete” tem uma grande chance de gerar dados inúteis.

Após o estabelecimento dos objetivos deve-se atentar para a delimitação e esclarecimento do que deve ser medido na avaliação, de forma clara e inequívoca. Apesar de ser desejável a realização de análises estatísticas *a posteriori* em busca de uma interpretação para as características instrucionais envolvidas, em termos de sua estrutura fatorial, acreditamos que as características devam possuir, *a priori*, um significado em termos dos objetivos. Necessitamos, portanto, antes de mais nada, da especificação das dimensões representativas do domínio de conhecimentos que está sendo estudado. Cada dimensão pode ser decomposta num conjunto de elementos que representam as características importantes a serem medidas, o que possibilitará um roteiro seguro para a geração dos itens que irão compor o instrumento. Assim, num instrumento destinado a avaliar prática docente, a dimensão de cumprimento de normas poderia ser decomposta, por exemplo, nos componentes pontualidade, assiduidade e devolução dos trabalhos e provas de acordo com os prazos regimentais.

Uma vez estabelecidos os objetivos e especificadas as dimensões, devemos considerar a incorporação das restrições como próximo passo a ser seguido na construção do instrumento. Esta etapa deve levar em consideração não só restrições de natureza temporal e orçamentária relativas ao instrumento em si (planejamento, construção, custos de elaboração e impressão, etc), mas deve ser visualizada como parte integral e harmoniosa do amplo conjunto de restrições da pesquisa como um todo. Como resultados desta fase temos a definição da quantidade de itens e o estabelecimento do tempo de aplicação do instrumento. A definição do formato e da modalidade de aplicação do instrumento, também características dessa fase, na maior parte das vezes, dependem de um julgamento subjetivo do pesquisador, levando em consideração todo o contexto da pesquisa. O formato preferencial, utilizado usualmente nos estudos sobre a avaliação pelo aluno, existente na literatura revista, tem sido o questionário impresso.

A escolha da escala de mensuração é uma decisão crucial na construção do instrumento, uma vez que a escala, por ser o veículo para a mensuração, em termos quantitativos, das características qualitativas das dimensões, pode ser considerada como o núcleo do instrumento. A literatura nos revela uma predominância por escalas do tipo

Likert na forma de itens do tipo discordo/concordo, com cinco ou sete opções de resposta para cada item.

Esta tipologia presta-se a uma interminável e inconclusiva discussão sobre o nível de mensuração subjacente, ordinal ou intervalar, bem como sobre os tipos de análises estatísticas que podem ser realizadas com os dados assim fornecidos (Babbie, 2003; Erthal, 2001; Hair, Babib, Money & Samouel, 2005; Lord, 1953; Michell, 1986; Nunnally & Bernstein, 1994; Pasquali, 1999, 2001, 2003; Stevens, citado por Lord, 1953; Townsend & Ashby, 1984). Além disso, o número de pontos da escala ou a existência, ou não, de um ponto neutro, é uma questão de preferência pessoal do pesquisador e não tem influência na estabilidade, validade e consistência interna da escala (Aiken, 1983; Bendig, 1954; Benson, 1971; Champney & Marshall, 1939; Cox, 1980; Garland, 1991; Hair & cols., 2005; Jacoby & Matell, 1971; Lissitz & Green, 1975; Matell & Jacoby, 1971, 1972; Pasquali, 1999, 2001, 2003).

O processo de seleção dos itens que irão compor o instrumento inicia-se pela geração de um conjunto abrangente de itens, sendo freqüentemente gerado mais de um item por elemento constituinte de cada dimensão, visando garantir uma maior cobertura do domínio de conhecimentos que é objeto da pesquisa. O número final de itens a ser incluído é determinado, usualmente, pelo julgamento do pesquisador, em função do contexto da pesquisa, e sua decisão pode ser suplementada pela análise estatística da consistência interna dos itens.

Para a geração do conjunto inicial de itens, poder-se-ia pensar, de início, na utilização de uma das seguintes estratégias: adota-se, na íntegra, um instrumento existente ou definem-se os itens, a partir do zero, com base nas dimensões do domínio de conhecimentos em estudo. A primeira estratégia é claramente inadequada, uma vez que dificilmente um instrumento utilizado numa pesquisa atende, na íntegra, aos objetivos de todas as pesquisas. A segunda estratégia, ao tentar reinventar a roda, torna-se também inadequada, pois não se beneficia de todo o conhecimento acumulado pela literatura. Uma estratégia promissora utiliza, à luz dos objetivos que se pretende atingir, duas abordagens complementares: na primeira selecionam-se itens já validados em instrumentos confiáveis, a partir de instrumentos ou banco de itens existentes na literatura, e na segunda criam-se novos itens gerados especificamente para a pesquisa a ser realizada.

A par das estratégias utilizadas, a seleção dos itens deve atender a pelo menos dois critérios de qualidade. Em primeiro lugar, os itens devem gerar boas respostas (Fowler,

1988). Para isso: a pergunta necessita ser comunicada e compreendida de forma consistente; as expectativas quanto à resposta adequada precisam ser claras para o respondente; a menos que se esteja verificando conhecimento, os respondentes devem ter toda a informação necessária para responder à pergunta; os respondentes precisam estar dispostos a responder (Fowler, 1988). Em segundo lugar, a linguagem utilizada na formulação dos itens do instrumento deve ser isenta de ambigüidade, compreensível à população-alvo da pesquisa e livre de viés e ênfase para não direcionar a resposta do respondente (Babie, 2003; Dillman, 2000; Günther, 2003; Malhotra, 2004; Pasquali, 1999, 2001, 2003).

A validação semântica dos itens representa uma análise teórica dos itens visando, essencialmente, evitar desperdício de recursos, devido à utilização de itens que não sejam adequados à população alvo do estudo. Ela representa uma síntese das etapas de teste piloto do instrumento, análise semântica dos itens e análise dos juízes (Pasquali, 1999).

Na ordenação dos itens devemos seguir uma estrutura lógica, agrupando os itens referentes a uma mesma dimensão ou aspecto num mesmo bloco e ordenando-os de acordo com um critério de relevância, iniciando-se o instrumento com aqueles blocos diretamente relacionados ao propósito do estudo. Esta orientação deve-se ao fato de que, em primeiro lugar, um questionário pode ser pensado como uma entrevista entre o pesquisador e o respondente, sendo a entrevista uma conversa com um objetivo (Bingham e Moore, citado por Günther, 2003). Em segundo lugar, responder a um questionário corresponde a uma conversa que flui de acordo com normas sociais: cumprimento, interação em si e uma despedida (Schwarz, citado por Dillman, 2000). Em terceiro lugar, a relevância do tópico a ser respondido é um dos principais preditores de taxa de resposta (Heberlein e Baumgartner, citado por Dillman, 2000). Em quarto lugar, um questionário deve ser interessante, curto, não exigir grandes esforços físicos e mentais para ser respondido e oferecer uma sensação de relevância e importância para as respostas oferecidas (Dillman, 2000).

Cabe esclarecer, ainda, que, em todos os momentos da construção do instrumento, devemos manter sempre a visão do estudo como um todo. Essa visão global faz com que os resultados das diversas etapas se inter-relacionem, gerando *feedbacks* e reformulações constantes, até a obtenção de uma forma do instrumento que seja julgada satisfatória pelo pesquisador.

## Capítulo 4 – Método

### 4.1 Locus da Pesquisa

O locus da pesquisa desenvolvida na presente tese foi uma Instituição de Educação Superior (IES) particular de Brasília, com larga tradição no cenário educacional local, e cuja missão, expressa em seu estatuto, é a de oferecer ensino superior de excelência à comunidade do Distrito Federal e entorno. Em maio de 2005, data da coleta de dados, a IES possuía 4.249 alunos matriculados em seus cinco cursos de graduação (Administração, Contabilidade, Direito, Economia e Pedagogia), com os cursos de Administração e Direito funcionando nos turnos matutino e noturno e os demais apenas no turno noturno.

O aluno típico da IES possuía o seguinte perfil:

- Solteiro (64%).
- Gênero feminino (51%).
- Faixa etária de 21 a 30 anos (59%).
- Renda familiar mensal superior a R\$5.000,00 (28%).
- Funcionário público (35%).
- Morador do Plano Piloto (35%).
- Utiliza transporte próprio para vir à IES (55%).
- Despende mais de 4 horas/semanais com a Internet (38%).
- Utiliza o computador principalmente para a realização de trabalhos escolares (38%).
- Pretende fazer pós-graduação (90%).
- Tem interesse em participar de atividades de Iniciação Científica (76%).

### 4.2 O Instrumento

A proposta metodológica de desenvolvimento de um instrumento de coleta de dados, ilustrada na Figura 3, foi viabilizada num questionário descrito a seguir. A escolha deste tipo de instrumento deveu-se a dois aspectos. O primeiro aspecto relaciona-se ao fato de que o questionário impresso é o instrumento preferencial utilizado na literatura relevante sobre a temática da avaliação do professor pelo aluno. O segundo aspecto relaciona-se à própria definição de questionário: um conjunto de perguntas sobre um determinado tópico

que não testa a habilidade do respondente, mas mede sua opinião, seus interesses, aspectos de personalidade e informação biográfica (Yaremko & cols., citado por Günther, 2003).

O questionário desenvolvido teve como objetivo a avaliação da prática docente no contexto da sala de aula. Para a consecução desse objetivo, foram utilizadas duas dimensões. A dimensão de prática de ensino foi decomposta nos elementos habilidades de comunicação, condução das aulas e coerência na proposição de avaliações. A dimensão de relacionamento professor-aluno foi considerada em sua totalidade, sem decomposição em elementos constituintes.

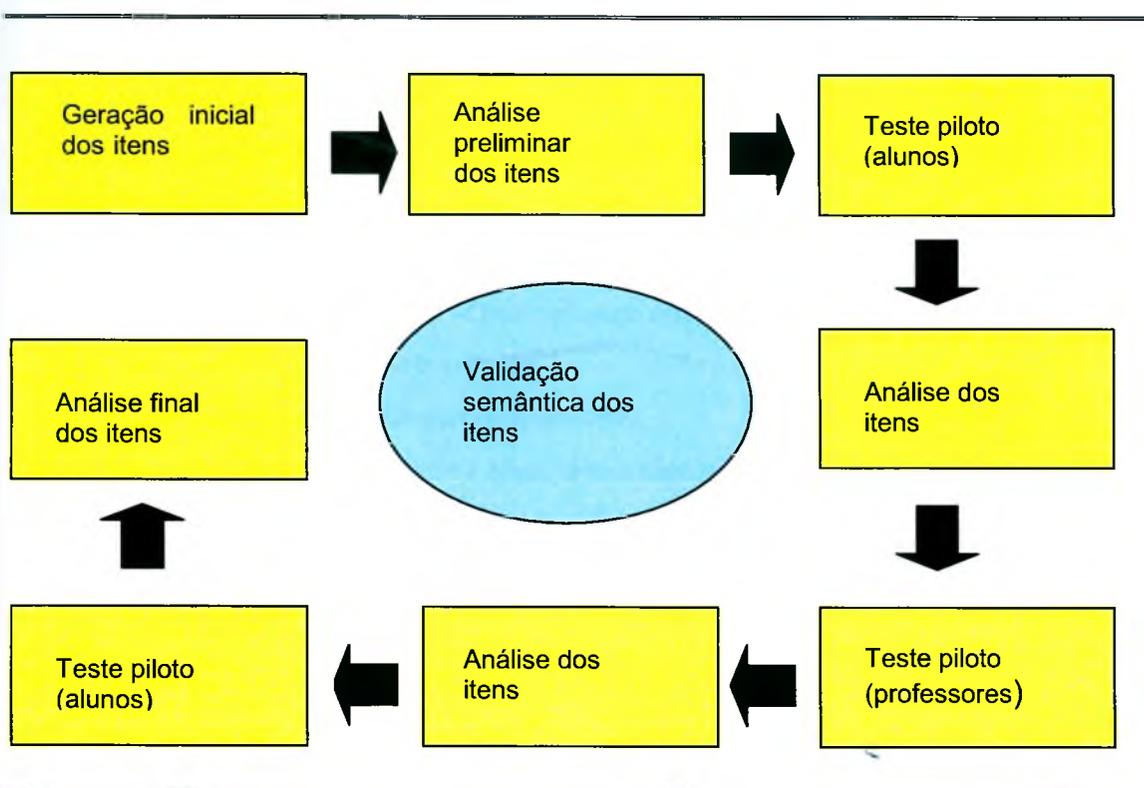
No caso da incorporação das restrições, levamos em conta o fato de que os alunos tendem a perder o interesse no processo de avaliação, caso este se mostre demorado ou muito prolixo ou caso a tarefa não seja breve ou fácil, ou seja aborrecedora ou aversiva (Scriven, 1995; Günther, 2003). Assim, o questionário planejado foi curto, tanto quanto ao número de questões quanto ao tempo de aplicação. O questionário utilizado foi preparado para ser lido por leitora óptica, impresso em duas cores na frente, contendo as questões fechadas, e uma cor no verso, contendo as questões abertas. Esta decisão levou em conta as restrições de natureza orçamentária, facilidade de manuseio, rapidez de apuração, minimização de erros de transcrição, redução de erros de mensuração, agradabilidade de leitura e boa taxa de resposta

No caso da escala de mensuração, utilizamos o formato presente na maioria dos instrumentos citados na literatura sobre avaliação do aluno pelo professor: uma escala do tipo Likert. Este tipo de escala é também um formato que não exige grandes esforços de resposta por parte dos alunos e se coaduna bem com as exigências de um questionário curto e simples. A escala utilizada foi uma escala de 10-pontos, variando de 1 a 10, com a faixa de 1 a 5 indicando uma avaliação ruim, a faixa de 6 a 8 caracterizando uma avaliação boa e a faixa de 9 a 10 representando uma avaliação ótima.

A escolha desta faixa deveu-se a três razões. A primeira razão é de natureza cultural, uma vez que a pontuação de 1 a 10 já é conhecida e utilizada pelos alunos das escolas brasileiras desde o ensino fundamental. A segunda razão é de natureza técnica, uma vez que evita a interminável e inclusiva discussão sobre o nível de mensuração subjacente às escalas do tipo Likert que utilizam itens ancorados na tipologia discordo/concordo. A terceira razão é de natureza política, uma vez que evitando-se a utilização do valor 0 (zero), diminui-se a resistência por parte dos professores que estão sendo avaliados.

Na seleção dos itens, além de atentarmos para os critérios de qualidade estabelecidos no Capítulo 3, utilizamos, de forma complementar, duas estratégias. Na primeira empregamos itens disponíveis em instrumentos já utilizados e validados na literatura (Aleamoni, 1978; Bem, 2004; Macedo, 2001; Marsh, 1982; Moreira, 1986, 1988; Pasquali, 1984; Santos, 2004). Na segunda desenvolvemos itens específicos para o alcance dos objetivos do estudo utilizando referências adicionais (Bain, 2004; Boclin, 2004; Cowan, 2002; Cunha, 1989; Lowman, 2004; Masetto, 2003; Miller, 1972; Millman, 1981; Morais, 1986, 1988; Morales, 2001; Moreira, 1988; Nikolic & Cabaj, 2001; Pereira & Morais, 2000; Pimenta & Anastasiou, 2002; Pimentel, 1993; Pinent, Silveira & Moraes, 1993; Rangel, 2002; Rios, 1993, 2002; Ryans, 1960; Sant'anna, 2002; Tardif, 2002; Tollefson, Cheng & Kleinsasser, 1989; Unesco, 2004).

A validação semântica dos itens foi realizado empregando-se o procedimento ilustrado na Figura 4.



**Figura 4.** Validação semântica dos itens do questionário.

O conjunto de itens foi submetido a uma análise inicial por parte de um comitê revisor, composto de seis professores da IES, com experiência em avaliação, e cujas características

encontram-se descritas no Anexo 1. O conjunto de itens revisado foi submetido a um teste piloto com 42 alunos dos cursos de Administração, Contabilidade, Direito, Economia e Pedagogia da IES. Os resultados do teste piloto foram submetidos a nova análise pelo comitê revisor. Um teste piloto com 200 professores dos cursos de Administração, Contabilidade, Direito, Economia e Pedagogia da IES foi aplicado ao conjunto de itens revisado. Os resultados do teste piloto foram analisados pelo comitê revisor. Um novo teste piloto com 35 alunos dos cursos de Administração, Contabilidade, Direito, Economia e Pedagogia da IES foi aplicado ao conjunto de itens revisado. Os resultados obtidos foram analisados pelo comitê revisor, que contou, nesta fase, com a participação de um professor de Língua Portuguesa, cujas características também se encontram descritas no Anexo 1. O conjunto final de itens acha-se descrito, de forma resumida, na Tabela 2.

**Tabela 2.** Itens do questionário.

---

O professor:

---

1. responde às perguntas e aos comentários dos alunos de forma educada e sem aparentar o irritação.
  2. promove a interação entre os alunos e encoraja o respeito mútuo.
  3. procura criar uma atmosfera agradável na sala de aula.
  4. mantém relacionamento cordial com os alunos.
  5. aplica as mesmas regras para todos, sem favorecer alguns alunos em relação a outros.
  6. resolve, com habilidade, os problemas de indisciplina em sala de aula.
  7. ministra as aulas com entusiasmo e bom humor.
  8. propõe tarefas interessantes para favorecer o aprendizado.
  9. usa uma linguagem clara para apresentar e desenvolver a matéria.
  10. responde aos questionamentos dos alunos de forma clara e objetiva.
  11. orienta, com clareza, os trabalhos propostos.
  12. desenvolve a matéria num ritmo satisfatório.
  13. encoraja os alunos a fazer questionamentos, expressando livremente suas idéias
  14. propõe provas que valorizam a reflexão mais do que a memorização
  15. demonstra coerência entre as avaliações e os conteúdos desenvolvidos em sala de aula.
  16. utiliza o tempo de aula de forma produtiva.
-

A prática docente acha-se distribuída, teoricamente, nos itens 1 a 7, referentes à dimensão de relacionamento professor-aluno, e nos itens 8 a 16, referentes à dimensão de prática de ensino. Todos esses itens foram respondidos na faixa numérica de 1 a 10.

O questionário completo está apresentado no Anexo 2. A ordenação dos itens foi feita de acordo com a orientação metodológica recomendada pelo modelo ilustrado na Figura 3. Na medida do possível encaixamos as questões imprescindíveis ao estudo na frente do questionário e as questões de natureza complementar no verso do questionário, pois a nossa experiência tem sugerido que há uma boa taxa de não resposta para questões que se encontram no verso do questionário. No verso do questionário foram incluídas duas questões abertas para que os alunos pudessem indicar tanto aspectos positivos do professor avaliado bem como aspectos em que o professor precisaria melhorar. No verso foi disponibilizado também um espaço para que o aluno oferecesse sugestões ou comentários. Os itens destinados à obtenção de dados referentes ao tamanho das turmas ou a características dos professores não precisaram ser incluídos no questionário, uma vez que eles puderam ser obtidos diretamente junto aos diversos setores da IES ao qual foi aplicado o questionário.

#### 4.3 Procedimentos

O período de aplicação dos questionários foi escolhido de forma a não coincidir com datas de provas, minimizando-se, na medida do possível, a influência das notas obtidas na avaliação dos professores pelos alunos. Apesar deste período estar devidamente incorporado ao calendário escolar da IES e já ser do conhecimento de toda a comunidade acadêmica, tivemos a preocupação de aumentar o esclarecimento sobre a avaliação e sensibilizar, tanto os alunos quanto os professores, para a importância do trabalho a ser realizado. Para isso, três dias antes do período previsto para o início da pesquisa, foram enviados e-mails, tanto para os alunos quanto para os professores. Estes e-mails acham-se descritos no Anexo 3.

No período de 16 a 21/05/2005 foi realizada a coleta dos dados. Cada aluno respondeu aos itens do questionário aplicado, avaliando, em cada horário normal de aula, o professor correspondente. O tempo médio sugerido para a aplicação do questionário foi de 15-20 minutos. A orientação para a aplicação dos questionários encontrava-se colada na parte de fora do envelope lacrado, retirado pelos professores no mesmo local em que assinam o

ponto e apanham as suas pautas de chamada. Esta orientação encontra-se descrita no Anexo 4.

#### 4.4 Participantes

Responderam aos questionários os estudantes presentes às aulas no período de 16 a 21/05/2005, representando alunos de 654 turmas e 227 professores. Os resultados obtidos acham-se descritos na Tabela 3.

**Tabela 3.** Respostas dos questionários por curso, sexo e turno  
(N=17.709)

curso	turno matutino			turno noturno			total
	gênero	gênero	não informado	gênero	gênero	não informado	
	feminino	masculino	não informado	feminino	masculino	não informado	
Administração	218	136	7	1.741	1.380	161	3.643
Contabilidade	-	-	-	1.115	1.086	129	2.330
Direito	741	531	82	3.254	4.151	602	9.361
Economia	-	-	-	338	461	43	842
Pedagogia	-	-	-	658	47	83	788
não informado	21	15	2	55	41	611	745
total	980	682	91	7.161	7.166	1.629	17.709

Foi obtido um total de 17.709 questionários respondidos, o que corresponde a 73,6% do universo possível de questionários passíveis de serem respondidos. Dentre os questionários respondidos, cerca de 90% foram respondidos por alunos do turno noturno e apenas 10% foram respondidos por alunos do turno matutino. O curso de Direito é responsável por mais da metade dos questionários respondidos (cerca de 53%), enquanto que as respostas dos alunos do curso de Pedagogia correspondem a cerca de 4% do total. As respostas dos alunos do gênero feminino correspondem a cerca de 46% do total de respostas fornecidas.

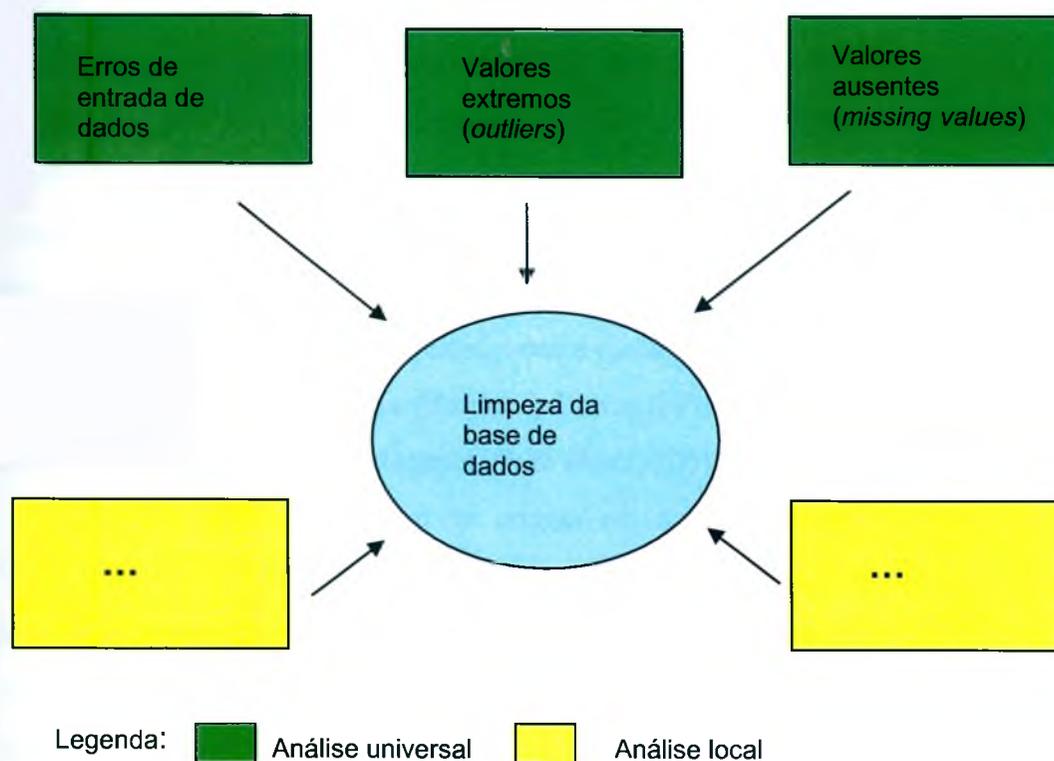
#### 4.5 Análise dos Dados

A análise dos dados foi feita empregando-se uma estratégia sequencial de três etapas. Na primeira etapa foi realizada uma análise exploratória dos dados, mediante a identificação e tratamento dos casos de erros de entrada de dados, valores extremos (*outliers*) e valores ausentes (*missing values*). Em seguida foi realizada uma análise preliminar da estrutura fatorial dos dados, juntamente com a análise de fidedignidade do

instrumento utilizado. Finalmente foi realizada uma análise confirmatória da estrutura fatorial dos dados.

Na análise preliminar foi utilizada a técnica de análise fatorial e na análise confirmatória foi empregada a técnica de modelagem por equações estruturais. A escolha dessas técnicas justifica-se por serem a análise fatorial e a modelagem por equações estruturais as técnicas de análise estatística multivariada recomendadas para análises estruturais (Afifi, Clark & May, 2004; Bartholomew, Steele, Moustaky & Galbraith, 2002; Bollen, 1989; Bryman & Cramer, 2003; Carvalho, 2004; Everitt & Dunn, 2001; Grimm & Yarnold, 2002, 2003; Hair & cols., 2005; Harlow, 2005; Kachigan, 1991; Laros, 2005; Lattin, Carrol & Green, 2003; Mingoti, 2005; Pasquali, 2006; Reis, 2001; Tacq, 1997; Ullman, 2001).

A análise exploratória dos dados, também conhecida como limpeza da base de dados, encontra-se ilustrada na Figura 5. As análises relativas a erros de entrada de dados, valores extremos e valores ausentes são universais, isto é, devem ser realizadas em qualquer base de dados, independentemente das análises estatísticas principais a serem realizadas. As demais análises são locais, isto é, são dependentes das análises estatísticas principais, uma vez que cada técnica estatística possui pressupostos específicos que devem ser atendidos pelos dados para que a técnica possa ser aplicada adequadamente.



**Figura 5.** Limpeza da base de dados.

A limpeza da base de dados, realizada após a coleta dos dados e antes das análises estatísticas principais a serem realizadas, é um procedimento que, apesar de tedioso e demorado, deve ser feito por qualquer pesquisador que deseje uma melhoria na acuidade dos dados. A limpeza da base de dados, além de possibilitar um conhecimento inicial das características dos dados empiricamente coletados, previne o *gargabe in, gargabe out* (GIGO), evitando assim que as análises estatísticas subsequentes sejam seriamente comprometidas ou totalmente desprovidas de sentido (Hartwig & Dearing, 1979; Laros, 2005; Pasquali, 2006; Tabachnick & Fidell, 2001).

Os erros de entrada de dados podem ser geralmente corrigidos pelo retorno aos documentos originais ou aos sujeitos da pesquisa. Um valor errado pode também ser corrigido mediante a sua substituição por um valor obtido à luz de informação pertinente à natureza do estudo.

Um valor extremo pode ser considerado como um elemento que se destaca no conjunto de dados de forma a ser considerado como duvidoso aos olhos do pesquisador

(Dixon, 1950), parecendo ter sido originado por um mecanismo diferente do que foi utilizado para os demais dados (Hawkins, 1980) ou mesmo como um elemento capaz de contaminar os demais dados (Wainer, 1976). Os valores extremos podem provocar uma inflação na taxa de erros e substancial distorção nas estimativas de parâmetros, tanto em procedimentos paramétricos quanto não-paramétricos (Zimmerman, citado por Osborne & Overbay, 2004). Os valores extremos podem, ainda, provocar um aumento da variância do erro, reduzir o poder dos testes estatísticos, alterar as probabilidades de cometimento dos erros do tipo I e tipo II nos testes estatísticos e violar os pressupostos de normalidade, tanto univariada quanto multivariada (Barnet & Lewis, 1994; Huck, 2000; Rassmussen, 1988; Schwager & Margolin, 1982; Tabachnick & Fidell, 2001).

Os valores extremos podem ter origem em diversas fontes (Osborne & Overbay, 2004). Em primeiro lugar, os valores extremos podem ser devidos a erros na entrada de dados. Em segundo lugar, os valores extremos podem se originar de respostas fornecidas com o caráter de desejabilidade social. Esta situação pode ocorrer em pesquisas que envolvem dados de natureza sensível, tais como renda, hábitos de higiene, experiência sexual, consumo de álcool ou drogas. Nestes casos os valores legítimos podem parecer valores extremos, quando, na realidade, não são. Em terceiro lugar, os valores extremos podem ser provenientes do método empregado na pesquisa, fazendo com que um grupo de respostas seja fornecido da mesma maneira, o que provoca o aparecimento de valores extremos. Em quarto lugar, os valores extremos podem ser provenientes da variabilidade natural dos dados. Devido à natureza das variáveis, podem ser obtidos dados que se apresentam como distribuições multimodais, assimétricas, etc. Estes dados, perfeitamente legítimos, podem ser interpretados, equivocadamente, pelo pesquisador como valores extremos, devido a erro no planejamento da pesquisa ou a suposições prévias inadequadas sobre a estrutura dos dados a serem obtidos. Em quinto lugar, os valores extremos podem ser devidos a procedimentos de amostragem, em que os valores obtidos foram selecionados de uma população diferente da população alvo do estudo.

Além da identificação dos valores extremos e suas possíveis origens, um aspecto importante da análise exploratória consiste no tratamento que deve ser dado a eles. No caso em que os valores extremos são valores incluídos de forma ilegítima parece haver consenso de que eles devem ser eliminados, uma vez que, além de inadequados, eles não são úteis para o estudo (Anscombe, 1960; Barnett & Lewis, 1994; Osborne & Overbay, 2004; Tabachnick & Fidell, 2001). Se os valores extremos são parte legítima dos dados ou se a

sua origem não é clara, não há consenso entre os pesquisadores, uma vez que há recomendações tanto para sua eliminação quanto para sua manutenção (Judd & McClelland, 1989; Orr, Sackett & DuBois, citado por Osborne & Overbay, 2004).

Uma das maneiras de se manter valores extremos legítimos é mediante o uso de transformações de variáveis, que trazem o benefício adicional de manter a relativa ordenação dos valores e eventualmente reduzir a variância do erro e a assimetria da distribuição (Hamilton, 1992; Osborne, 2002; Tabachnick & Fidell, 2001). O uso de transformações de variáveis, no entanto, ao mudar a métrica de mensuração, pode fazer com a transformação se torne inadequada para a interpretação do modelo em estudo ou dificulte a sua interpretação (Newton & Rudestam, 1999; Osborne, 2002). A utilização de procedimentos robustos é uma outra maneira de se proteger os dados da influência de valores extremos (Anscombe, 1960; Davies & Gather, 1993; Ludbrook & Dudley, 1998; Rousseeuw & Leroy, 1987; Wilcox, 1992, 1993, 1994, 1996, 1997, 2003).

Valores ausentes são um dos problemas mais pervasivos na análise de dados (Tabachnick & Fidell, 2001). Valores ausentes podem não só reduzir a precisão de estatísticas uma vez que há menos informação que a que foi originalmente planejada para a obtenção dessas estatísticas, bem como comprometer os resultados das análises (Hair & cols., 2005; Pasquali, 2006). O tratamento para os dados ausentes consiste na sua substituição ou eliminação (Rubin, 1987; Little & Rubin, 1987).

Entre as alternativas de substituição temos escolhas *ad hoc*, tais como média, regressão ou combinação de padrões, sendo que a substituição dos valores ausentes representa, na melhor das hipóteses, uma escolha entre diversas alternativas ruins (Tabachnick & Fidell, 2001). A eliminação dos valores ausentes pode ser feita pelos métodos *pairwise* ou *listwise*. Na aplicação do método *pairwise*, somente casos possuindo valores ausentes nas variáveis indicadas para um determinado cálculo são excluídos da análise. O método *listwise* elimina todos os casos em que ocorre valores ausentes em qualquer das variáveis. Este método tem o inconveniente, em geral, de reduzir bastante o tamanho da amostra a ser utilizada nas análises, o que faz com que ele seja recomendado apenas para os casos em que o número de valores ausentes não seja muito grande (Byrne, 2001). Esta recomendação não é de muita utilidade, uma vez que não existem regras ou diretrizes claras sobre o que seja um número não muito grande de valores ausentes, apesar da existência de uma regra empírica que sugere que este número não deve ultrapassar 10% do tamanho da amostra (Kline, 1998).

Como vimos, a não realização da limpeza da base pode ter influência devastadora nos resultados de uma pesquisa. Apesar desse perigo, as recomendações, listadas nesta tese, para o tratamento de valores extremos e valores ausentes devem ser encaradas não como camisa de força mas apenas como sugestões que podem auxiliar o pesquisador a tomar boas decisões.

A segunda etapa da análise realizada neste trabalho foi a análise preliminar da estrutura dos dados mediante a técnica de análise fatorial. A análise fatorial é uma técnica imprescindível no contexto da Psicometria, sobretudo para a problemática da validação de instrumentos psicológicos, bem como da Psicologia em geral e das ciências psicossociais em particular (Pasquali, 2005). Ela é empregada largamente na avaliação de construtos psicológicos (Laros, 2005).

A análise fatorial pode ser empregada para analisar inter-relações entre um grande número de variáveis e explicar essas variáveis em termos de suas dimensões inerentemente comuns ou seja ela permite condensar a informação contida em um número de variáveis originais em um conjunto menor de variáveis estatísticas com perda mínima de informação (Hair & cols., 2005). Numa visão explicativa, a análise fatorial é empregada de forma preliminar com a finalidade de identificar as dimensões subjacentes que representam os construtos teóricos do instrumento. Assim, numa visão explicativa, a finalidade da análise exploratória é a descoberta das variáveis latentes que estão subjacentes à escala (Laros, 2005).

A análise fatorial fornece três resultados principais (Laros, 2005). Em primeiro lugar, ela produz cargas fatoriais que podem ser consideradas como pesos de regressão das variáveis mensuradas para predizer o construto subjacente. Em segundo lugar, caso haja mais de um fator subjacente aos dados, ela fornece as correlações entre os fatores. Em terceiro lugar ela separa a variância de cada variável mensurada em duas partes: compartilhada e única. A variância compartilhada (comunalidade) é a variância associada às variáveis latentes, estimada com base na variância partilhada com outras variáveis mensuradas na análise. A variância única é a combinação da variância específica para a variável mensurada com a variância do erro aleatório na mensuração da variável. A variância verdadeira, representada pela soma da comunalidade com a variância específica (sem o erro), define a precisão e a variância compartilhada define a validade (Pasquali, 2005).

O outro uso da análise fatorial está relacionado à redução de dados. O objetivo da redução de dados é descobrir ponderações ótimas para as variáveis mensuradas, de forma que um grande conjunto de variáveis possa ser reduzido a um conjunto menor de índices sumários que tenham máxima variabilidade e fidedignidade (Laros, 2005). Vê-se, portanto, que a análise fatorial está apoiada no princípio da parcimônia, uma vez que fornece um número menor de variáveis (fatores) como suficiente para explicar uma série maior de variáveis observáveis (Pasquali, 2005).

Para que a análise fatorial seja realizada a contento é importante que o pesquisador esteja atento a algumas considerações, sendo a primeira delas relacionada à natureza da amostra. Uma vez que quanto mais heterogênea uma amostra, em relação às variáveis que estão sendo mensuradas, mais altas são as correlações encontradas, é recomendado para a análise fatorial a utilização de amostras apropriadamente heterogêneas que representem toda a extensão da população alvo (Clark & Watson, 1995; Laros, 2005).

Uma segunda consideração que o pesquisador deve levar em conta relaciona-se ao tamanho da amostra. Sobre o tamanho mínimo da amostra a ser utilizada na análise fatorial são feitas algumas recomendações. Uma primeira recomendação é a de considerar cinco participantes por variável e uma amostra total de pelos menos 200 sujeitos (Gorsuch, 1983). Uma segunda recomendação é a de considerar dez sujeitos por variável, com um mínimo de 100 sujeitos na amostra total (Crocker & Algina, 1986). Uma terceira recomendação é a de um mínimo de 100 sujeitos por fator medido (Pasquali, 2005). Além dessas regras há a sugestão de se classificar uma amostra de tamanho 50 como muito inferior, uma amostra de tamanho 100 como inferior, uma amostra de tamanho 200 como razoável, uma amostra de tamanho 300 como boa, uma amostra de tamanho 500 como muito boa e uma amostra de pelo menos 1.000 sujeitos como excelente (Comrey & Lee, 1992).

Há, no entanto, pesquisadores que argumentam que não há base teórica ou empírica para a recomendação de relação entre o número de participantes e o número de variáveis (Gaudagnoli & Velicer, 1988) ou que não existe um tamanho mínimo de amostra para se efetuar uma análise fatorial com um determinado número de fatores (Wolins, 1995). Estas últimas considerações parecem ser as mais pertinentes, uma vez que estudos mais recentes revelaram que o tamanho adequado da amostra é parcialmente determinado pela natureza dos dados (Fabrigar, Wegener, MacCallum & Strahan, 1999; MacCallum, Widaman, Zang & Hong, 1999). Os resultados obtidos nestes estudos indicam que quanto mais fortes os

dados, menor a necessidade do tamanho da amostra para a obtenção de uma análise acurada.

Como na teoria as variáveis latentes são as causas subjacentes das variáveis mensuradas, a análise fatorial utiliza as correlações entre estas variáveis. Assim, uma terceira consideração que deve ser levada em conta refere-se à natureza e escala de mensuração das variáveis, uma vez que estes aspectos são determinantes no procedimento correto a ser adotado (O'Connor, 2006). Para este autor, as matrizes de correlações a serem usadas deveriam ser a matriz de correlações de Pearson para variáveis contínuas, a matriz de correlações tetracóricas para dados dicotômicos e a matriz de correlações policóricas para dados ordinais.

Uma quarta consideração a ser levada em conta refere-se às distribuições das variáveis, uma vez que a correlação entre duas variáveis é afetada tanto pela similaridade entre sua substantividade quanto pela similaridade entre suas distribuições estatísticas (Bernstein, Garbin & Teng, 1988). Assim, variáveis com distribuições similares tendem a se correlacionar de maneira mais forte umas com as outras do que com variáveis com distribuições não similares e, em consequência, a formação de fatores pode ser decorrente apenas da similaridade das distribuições das variáveis. Para se verificar este efeito basta computar as médias e os desvios-padrão de cada uma das variáveis componentes dos diversos fatores. Se os resultados mostram um fator com a maioria dos itens com altos valores, outro com médios valores e outro com baixos valores, existe uma forte razão para atribuir os fatores a bases estatísticas e não a bases substantivas (Bernstein, Garbin & Teng, 1988).

Uma vez atendidos os pressupostos anteriormente descritos deve ser realizada a análise fatorial propriamente dita. Esta realização é feita mediante as seguintes etapas: verificação da fatorabilidade da matriz de correlações, determinação do número de fatores a extrair, escolha dos métodos de extração e rotação dos fatores, interpretabilidade e refinamento da estrutura fatorial (Hair & cols., 2005; Laros, 2005; Laros & Puente-Palacios, 2004; Pasquali, 2005).

O primeiro passo na análise da estrutura fatorial do instrumento consiste no exame da fatorabilidade da matriz de correlações entre os itens, uma vez que se eles não têm variância em comum, a análise fatorial torna-se sem sentido (Laros & Puente-Palacios, 2004). Este exame consiste na inspeção da matriz de correlações, no cálculo do índice de adequação da amostra de Kaiser-Meyer-Olkin (KMO) e no teste de esfericidade de Bartlett.

Na inspeção da matriz de correlações, a sua fatorabilidade é sugerida pela existência de correlações altas fora da diagonal principal e de valores nulos na parte da significância (Pasquali, 2005). Se o determinante dessa matriz for zero tem-se um único fator, caso contrário tem-se a indicação de mais de um fator (Pasquali, 2005). O índice KMO está baseado nas correlações parciais entre as variáveis e quanto mais próximo de 1, maior a comunalidade entre as mesmas e, portanto, melhor a fatorabilidade do conjunto (Laros & Puentes-Palacios, 2004). Valores a partir de 0,6 para o KMO podem ser consideráveis toleráveis (Pasquali, 2005). O teste de esfericidade de Bartlett testa a hipótese nula de que a matriz de correlações é uma matriz identidade, ou seja, as correlações entre as variáveis é nula. Para que a análise fatorial faça sentido, a hipótese nula do teste deve ser rejeitada (Laros, 2005).

Uma vez admitida a fatorabilidade da matriz de correlações, o próximo passo da análise fatorial consiste na determinação do número de fatores a serem extraídos. A determinação do número correto de fatores a serem retidos na análise fatorial é um passo crucial na construção de instrumentos psicométricos e na elaboração de teorias psicométricas, uma vez que a ocorrência de erros nesta fase afetará a interpretação de todos os resultados subsequentes (Laros, 2005). Assim, a determinação do número de fatores a extrair é provavelmente a decisão mais importante que um pesquisador, executando análise fatorial, tomará (Zwick & Velicer, 1986). Desta forma a superextração ou subextração de fatores levarão a resultados e conclusões distorcidas (Fava & Velicer, 1996; Linn, 1968; Velicer & Jackson, 1990). Os critérios mais comumente empregados para a determinação dos fatores são: autovalor maior que 1, qui-quadrado, inspeção do *scree-plot*, correlações parciais, análise paralela e percentual de variância explicada.

O critério do autovalor maior que 1 (Guttman, 1954; Kaiser, 1960), também conhecido como critério de Kaiser-Guttman ou simplesmente K1 ou ainda GK, é um critério bastante popular em Psicologia e ciências sociais e é a opção padrão no software SPSS. Uma vez que o autovalor representa a quantidade da variância explicada por um fator, um valor 1 para o autovalor representa a porcentagem da variância explicada por uma única variável. Assim o critério GK é baseado na consideração de que um fator precisa explicar pelo menos a quantidade de variância que é explicada por uma única variável (Laros, 2005; Pasquali, 2005; Tabachnick & Fidell, 2001).

O critério do qui-quadrado é baseado no teste de qui-quadrado (Bartlett, 1950), que testa a hipótese nula de que os autovalores de todos os fatores são iguais. Começando com

o autovalor do último fator, cada autovalor é excluído sequencialmente até que a hipótese nula do teste seja rejeitada. Este teste parece ser sensível ao tamanho da amostra, levando à retenção de mais fatores em amostras grandes (Gorsuch, 1983). O teste do qui-quadrado de Bartlett não está disponível no software SPSS.

*Scree* é um termo da Geologia e significa o cascalho ou pedregulho que se espalha no sopé da montanha (Pasquali, 2005). O critério da inspeção visual do *scree-plot* (Cattell, 1966), é baseado em um gráfico dos autovalores: no eixo horizontal tem-se o número sequencial dos autovalores e no eixo vertical os valores dos autovalores. Por este critério, a determinação do número de fatores baseia-se na inspeção visual do gráfico, buscando-se um ponto de quebra na curvatura natural do gráfico, isto é, um ponto em que o gráfico passa de uma inclinação acentuada para uma inclinação quase horizontal. O número de pontos à esquerda do ponto de quebra é o número de fatores a reter. Tal inspeção, subjetiva, pode não ter sucesso, uma vez que os valores dos autovalores podem mostrar um declive gradual com nenhum ponto de quebra claramente marcado ou mesmo apresentar mais de um ponto de quebra (Laros, 2005). O critério da inspeção visual do *scree-plot* apresenta uma acuidade maior em amostras grandes e fatores fortes (Zwick & Velicer, 1982).

O critério de correlações parciais (Velicer, 1976), também conhecido como critério MMCP de Velicer, é baseado na matriz de correlações parciais. Por este critério determina-se a matriz de correlações parciais após a extração de cada um dos fatores, calcula-se o quadrado das médias destas correlações e interrompe-se o processo quando este valor atinge um mínimo (Laros, 2005).

O método de análise paralela (Horn, 1965), compara os autovalores empíricos com os autovalores obtidos por meio de matrizes geradas aleatoriamente e que contêm variáveis aleatórias não-correlacionadas, usando-se tamanhos de amostras iguais ao da matriz de correlação empírica. Pela sugestão desse autor, seriam geradas aproximadamente 50 matrizes aleatórias, realizadas análises fatoriais nestas matrizes e calculadas as médias dos autovalores obtidos. Seriam extraídos somente fatores que correspondem a autovalores empíricos que excedem os valores médios dos autovalores obtidos de forma aleatória e os demais seriam considerados como devidos à variância amostral aleatória. Desta forma, a análise paralela, ao comparar os autovalores, tem o seu foco dirigido para a magnitude de variância explicada pelos fatores (Laros, 2005). Assim faz sentido a retenção de um fator

apenas na medida em que este explica maior variância do que o fator correspondente nos dados aleatórios (Reise, Waller & Comrey, 2000).

Os méritos relativos dos diferentes critérios para a determinação do número de fatores a serem extraídos têm sido objeto de diversos estudos, envolvendo principalmente dados simulados. Mostrou-se que o método GK resulta em severa superextração de fatores em comparação com o método de análise paralela (Horn, 1965). Num estudo envolvendo sete fatores predeterminados, com 20 e 40 variáveis e amostras de 100 e 50 sujeitos, mostrou-se que o método GK superestimou o número de fatores corretos em 66% (Linn, 1968). Num estudo envolvendo matrizes de correlações com três e sete fatores, 20 e 40 variáveis, amostras de 1.000 e 500 sujeitos e entre 40 a 50 replicações por condição, mostrou-se que o critério de análise paralela estava quase sempre correto (Humphreys & Montanelli, 1975).

Um outro estudo envolveu dois níveis fatoriais com seis e doze fatores conhecidos, dois níveis de saturação nos fatores com cargas fatoriais de 0,50 e 0,80 (Zwick & Velicer, 1986). Neste estudo foram construídas matrizes de correlação com 36 e 72 variáveis, tamanhos de amostras diferentes (com 2 e 5 vezes mais sujeitos que variáveis) e seis níveis de complexidade fatorial diferente, que variavam de uma estrutura fatorial simples até estruturas muito complexas. O critério GK indicou o número correto de fatores em apenas 22% das vezes e superestimou o número de fatores todas as vezes em que indicou o número de fatores errados. O critério do qui-quadrado mostrou-se correto 30% das vezes, com uma tendência a superestimar o número de fatores com o tamanho crescente da amostra. O critério de inspeção visual do *scree-plot* mostrou 57% de precisão e superestimou o número correto de fatores em 90% dos casos em que estava incorreto. O critério MMCP foi preciso em 84% dos casos e mostrou uma tendência para subestimação dos fatores em 90% dos casos quando estava errado. O método de análise paralela foi preciso em 92% dos casos e mostrou uma leve tendência para superestimar o verdadeiro número de fatores em 66% dos casos.

Outros estudos fornecem um amplo apoio para o uso da análise paralela em detrimento dos outros critérios (Crawford & Koopman, 1973; Franklin & cols., 1995; Hayton, Allen & Scarpello, 2004; Hubbard & Allan, 1987 e Humphreys & Ilgen, 1969). A análise paralela foi disponibilizada inicialmente no software RanEigen (Enzmann, 1977), que possui uma limitação máxima de 25 variáveis e 8.000 casos. A utilização da análise paralela no SPSS só foi possível com o desenvolvimento de uma rotina (sintaxe) que consta do trabalho de O'Connor (2000).

Pode-se utilizar também o critério da variância explicada para se decidir pelo número de fatores. O desejável é que os fatores extraídos expliquem o maior percentual possível de variância explicada (Figueiredo, 2001). Na literatura não há concordância quanto ao referencial a ser utilizado (Hattie, 1985). Assim, para Gorsuch (1983), o primeiro fator deve explicar um mínimo de 70% da variância total, para Streiner (citado por Hattie, 1985) o valor deve ser 50%, para Carmine e Zeller (1979) o valor deve ser 40% e para Reckase (citado por Hattie, 1985) o valor deve ser 20%. Um percentual de 30% é considerado tolerável (Pasquali, 2005). Levando-se em conta o critério da importância do fator, o valor mínimo de explicação da variância para cada fator deve ser 3% (Harman, 1976).

Na determinação do número de fatores a ser extraído um auxílio pode ser prestado pela matriz residual de correlações. Esta matriz resulta da diferença entre a matriz de correlações empíricas e a matriz de correlações produzidas a partir da matriz fatorial. Se a análise fatorial tiver sido adequada, a matriz residual deve ser composta de correlações próximas de zero, sendo que resíduos entre 0,05 e 0,10 são considerados de moderados a grandes (Tabachnick & Fidell, 2001). Se a matriz residual apresentar mais que 5% destes resíduos tem-se a indicação de que há mais fatores a serem extraídos (Pasquali, 2005).

Após a decisão sobre o número de fatores a ser extraído, passa-se ao método de extração de fatores. Nesta situação três métodos disputam a preferência dos pesquisadores: análise de componentes principais (PCA), análise de fatores principais (PFA) e máxima verossimilhança (ML).

O método PCA é um método que se tornou popular desde o tempo em que os computadores eram lentos e caros (Costello & Osborne, 2005). Apesar de sua enorme popularidade e de ser a opção padrão no software SPSS, sua utilização como método de análise fatorial é objeto de discordância entre os pesquisadores.

De um lado, há aqueles que defendem o ponto de vista de que não há diferenças entre PCA e outros métodos em análise fatorial ou que PCA deve ser preferido (Arrindel & van der Ende, 1985; Schonemann, citado por Costello & Osborne, 2005; Steiger, citado por Costello & Osborne, 2005; Velicer & Jackson, 1990). De outro lado, há aqueles que argumentam que o método PCA é simplesmente um método de redução de dados e não um método de análise fatorial (Bentler & Kano, 1990; Floyd & Widaman, 1995; Gorsuch, 1990; Loehlin, 1990; MacCallum & Tucker, 1991; Snook & Gorsuch, 1989; Widaman, 1990, 1993).

O método PCA não faz pressuposições sobre a estrutura subjacente às variáveis latentes e os fatores (componentes) são calculados usando toda a variância das variáveis mensuradas e toda a variância aparece na solução, uma vez que o método não discrimina entre a variância comum e a variância única. Isto contraria o espírito da análise fatorial no sentido de que um de seus resultados desejados é o particionamento da variância total. Além disso, quando os fatores são não correlacionados e a comunalidade é moderada, o método PCA pode produzir valores exagerados para a variância explicada pelos componentes (McArdle, 1990).

O método ML é o método que deve ser empregado, caso a distribuição dos dados seja normal multivariada ou aproximadamente normal multivariada (Costello & Osborne, 2005). Ele permite o cálculo de um grande número de índices de adequação do modelo, o cálculo de intervalos de confiança e a realização de testes de significância com as cargas fatoriais e as correlações entre os fatores (Fabrigar & cols., 1999). Caso o pressuposto de distribuição normal multivariada seja severamente violado deve ser usado um método de fatores principais (Fabrigar & cols., 1999). No SPSS o método *principal axis factoring* (PAF) é um método de análise de fatores principais.

Após a extração pode-se utilizar um método de rotação dos fatores visando simplificar e melhorar a interpretabilidade da estrutura fatorial. Os métodos de rotação podem ser divididos em duas grandes categorias: ortogonal e oblíqua. A rotação ortogonal produz fatores independentes, não correlacionados, enquanto que a rotação oblíqua permite a ocorrência de fatores correlacionados. Esta última categoria é prevalente em Psicologia e em ciências sociais, uma vez que nestas áreas do conhecimento geralmente espera-se algum grau de correlação entre os fatores (Pasquali, 2005).

Após a extração e rotação dos fatores deve ser buscada uma eventual simplificação e melhoria na interpretabilidade da estrutura fatorial obtida, uma vez que a maioria das estruturas fatoriais não é uma estrutura simples, ou seja, uma estrutura em que existe uma única carga fatorial alta para cada variável em um único fator (Hair & cols., 2005). Do ponto de vista substantivo, a interpretabilidade da estrutura fatorial deve levar em conta que o fator deve ter significado na teoria de interesse do pesquisador, isto é, o fator deve poder ser interpretado em termos de hipóteses pertinentes ao seu campo de pesquisa (Pasquali, 2005). Assim, tão importante quanto investigar as características psicométricas do item é verificar se a estrutura consegue representar, de maneira acurada, o conteúdo abordado pelo fator (Smith & McCarthy, 1995). Portanto, um fator para ser uma medida

válida deve possuir itens válidos e ser ele próprio válido, isto é, possuir itens com cargas fatoriais elevadas e em número suficiente para cobrir os aspectos do construto teórico abrangido pelo fator (Pasquali, 2005).

Uma estrutura fatorial pode ser considerada adequada se cada fator possui um número razoável de itens, se os itens possuem uma carga fatorial mínima e se há poucos itens com cargas em mais de um fator (Costello & Osborne, 2005). Na busca desta estrutura adequada são sugeridas algumas regras empíricas. A primeira delas estabelece 0,32 como o mínimo aceitável para a carga fatorial de um item (Tabachnick & Fidell, 2001). Esta regra encontra eco no fato de que este valor representa aproximadamente 10% de variância compartilhada com os outros itens no fator (Pasquali, 2005). Uma segunda regra empírica considera uma carga fatorial de no mínimo 0,71 como excelente, uma carga fatorial de no mínimo 0,63 é considerada muito boa, uma carga fatorial de no mínimo 0,55 é considerada boa, uma carga fatorial de no mínimo 0,45 é considerada razoável e uma carga fatorial abaixo de 0,32 é considerada pobre (Comrey, citado por Pasquali, 2005). Uma terceira regra empírica estabelece que um fator com menos de três itens é geralmente fraco e instável, enquanto que um fator com pelo menos 5 itens com cargas fatoriais mínimas de 0,50 é desejável e indica um fator sólido (Costello & Osborne, 2005).

Para se verificar a estabilidade do fator pode-se recorrer à análise da fidedignidade (Pasquali, 2005). A fidedignidade pode ser obtida com o auxílio da consistência interna. A idéia da consistência interna é que os itens individuais da escala devem medir o mesmo construto e assim ser altamente correlacionados (Nunnally & Bernstein, 1994). A avaliação da consistência interna pode ser obtida por meio da correlação item-total ou da correlação inter-itens, havendo regras empíricas que sugerem que a correlação item-total deve ser superior a 0,50 enquanto que a correlação inter-itens deve ser superior a 0,30 (Hair & cols., 2005). A avaliação da consistência da escala como um todo pode ser obtida pelo coeficiente alfa de Cronbach (Cronbach, 1951). Regras empíricas sugerem que valores deste coeficiente superiores a 0,90 representam um bom nível de consistência interna, sendo que em muitos casos um valor mínimo de 0,70 é considerado razoável (Hair & cols., 2005; Hogan, 2006; Lattin, Carrol & Green, 2003; Nunnally & Bernstein, 1994). Valores desse coeficiente em torno de 0,80 são considerados como aceitáveis (Pasquali, 2003).

A terceira e última etapa da análise realizada neste trabalho foi a análise confirmatória da estrutura fatorial dos dados mediante a técnica de modelagem por equações estruturais. Esta técnica permite ao pesquisador testar estruturas fatoriais previamente estabelecidas,

assim como analisar relações explicativas entre múltiplas variáveis simultaneamente, o que faz com ela possa ser entendida como uma mistura de análise fatorial e análise de caminhos (Bollen, 1989; Harlow, 2005; Kachigan, 1991; Kaplan, 2000; Kline, 1998; Lattin, Carrol & Green, 2003; McDonald & Ho, 2002; Ullman, 2001).

A modelagem por equações estruturais é uma técnica que exige, em primeiro lugar, amostra grande (Byrne, 2001). Esta exigência pode ser traduzida numa regra empírica que estabelece um mínimo de dez sujeitos para cada variável medida para que se possa utilizar a modelagem por equações estruturais (Mueller, citado por Pasquali, 2006). Em segundo lugar a modelagem por equações estruturais exige valores completos (Byrne, 2001). Esta exigência prende-se ao fato de que a existência de valores ausentes pode acarretar a não convergência para uma solução (Byrne, 2001). Apesar de alguns métodos de estimação exigirem que a distribuição das variáveis seja normal multivariada, alguns estudos revelaram que as estimativas dos parâmetros permanecem válidas sob determinadas condições mesmo quando os dados não são distribuídos normalmente (Amemiya & Anderson, 1990; Anderson, 1989; Browne & Shapiro, 1988; Satorra & Bentler, citado por McDonald & Ho, 2002).

A modelagem por equações estruturais representa uma área em plena atividade e com um oferecimento bastante grande de software: AMOS, CALIS, COSAN, EQS, LISREL, Mx, MPLUS, RAMONA e SEPATH. Uma vez que diferentes formulações matriciais de um modelo de caminhos com variáveis latentes são essencialmente equivalentes, esses programas computacionais fornecem as mesmas informações básicas, a menos de pequenas variações nos detalhes de implementação (McArdle & McDonald, 1984).

O caráter confirmatório da técnica reside no fato de que o modelo geral da modelagem por equações estruturais procura replicar um conjunto de dados observados por meio da imposição de parâmetros nas matrizes de covariâncias representativas das relações teóricas definidas pelo pesquisador. (Bollen, 1989; Byrne, 2001; Harlow, 2005; Kachigan, 1991; Kaplan, 2000; Kline, 1998; McDonald & Ho, 2002). O modelo hipotetizado pode, então, ser testado numa análise simultânea do sistema inteiro de variáveis para determinar a extensão da sua consistência com os dados. Se a consistência é adequada o modelo pode ser confirmado como plausível para a relação hipotetizada entre as variáveis.

Uma vez que a consistência pode ser entendida sob diversos pontos de vista, existem diversos índices de adequação do modelo, sendo que alguns são específicos do método de estimação usado (Arbuckle & Wothke, 1999, Byrne, 2001; Pasquali, 2006). A área de

desenvolvimento de índices de adequação é uma área em pleno vigor com os pesquisadores criando índices com base nos mais diferentes critérios (Boomsma, 2000; McDonald & Ho, 2002; Ullman, 2001). Somente o programa AMOS 4.0 oferece 31 diferentes tipos de índices de adequação (Arbuckle & Wothke, 1999).

O problema do uso dos índices de adequação está longe de obter consenso, sendo reconhecidos quatro problemas básicos com eles (McDonald & Ho, 2002). Primeiro, não há uma base empírica ou matemática para o uso dos índices de adequação. Segundo, não existe fundamento sólido que justifique a escolha de um índice relativo em relação a um índice absoluto ou para considerar variáveis não correlacionadas como um modelo vazio. Terceiro, não existe uma forte e suficiente correspondência entre índices alternativos que permitam tomar uma decisão baseada em um ou em outro. Quarto, e talvez o mais importante, um dado grau de inadequação global pode ser devido a uma má especificação que pode ser corrigida, por se dever a poucas discrepâncias, ou pode ser devida a uma dispersão geral de discrepâncias não associadas com qualquer má especificação em particular.

Apesar da aparente falta de consenso, três índices são bastante populares nas análises envolvendo modelagem por equações estruturais: *comparative fit index*, *goodness-of-fit index* e *root mean square error of approximation* (Pasquali, 2006). O índice *comparative fit index* (CFI) compara o modelo em teste com um modelo base, que admite que todas as variáveis do modelo são não correlacionadas. Este índice varia de 0 a 1 e quanto mais próximo de 1 melhor a adequação do modelo (Pasquali, 2006). O índice *goodness-of-fit index* (GFI) foi criado para os métodos de estimação de máxima verossimilhança e mínimos quadrados não ponderados e sua interpretação é semelhante ao índice CFI (Pasquali, 2006). O índice *root mean square error of approximation* (RMSEA) está baseado na avaliação dos resíduos que resultam da diferença entre a matriz original de variância-covariância e a matriz reproduzida. Resultados de pesquisa sugerem que um valor de cerca de 0,05 ou menos representa um valor mais realista que o índice ideal de 0 (Browne & Cudeck, citado por Arbuckle & Wothke, 1999). Uma regra empírica estabelece que um valor para o RMSEA menor que ou igual a 0,08 indica suficiente adequação do modelo, enquanto que um valor acima de 0,1 recomenda a não utilização do modelo (Browne & Cudeck, citado por Arbuckle & Wothke, 1999; Pasquali, 2006).

## Capítulo 5 – Resultados e Discussão

### 5.1 Análise Exploratória

No caso específico desta tese, pelo procedimento adotado, de leitura óptica dos questionários, na entrada de dados poderiam ter ocorrido erros na leitura dos questionários ou erros de múltipla marcação de respostas. Para se evitar o erro de leitura dos questionários, foi feita, previamente, uma inspeção visual, manual, de todos os questionários preenchidos e estabelecido um programa de controle para se evitar a não leitura de algum questionário ou a leitura em multiplicidade de um mesmo questionário. Assim, os erros de entrada de dados foram reduzidos a erros devidos a múltipla marcação de respostas para um mesmo item do questionário. A leitura óptica dos questionários aplicados gerou uma base com 17.709 observações. Na Tabela 4 apresentamos a frequência de múltipla marcação de respostas para os diversos itens do questionário.

**Tabela 4.** Ocorrência de múltipla marcação de respostas para os itens do questionário (N= 17.709).

item	freqüência absoluta	freqüência relativa
16	67	0,38 %
9	74	0,42 %
8	79	0,45 %
11	79	0,45 %
15	82	0,46 %
10	83	0,47 %
14	91	0,51 %
12	93	0,53 %
13	93	0,53 %
7	98	0,55 %
6	121	0,68 %
2	124	0,70 %
3	126	0,71 %
5	129	0,73 %
1	140	0,79 %
4	175	0,99 %

A Tabela 4 nos revela que o percentual de múltipla marcação de respostas variou de 0,38% a 0,99%. Assim, apesar do intervalo de variação de valores poder ser considerado razoalmente grande, tem-se um percentual muito baixo de múltipla marcação de respostas,

considerando cada item isoladamente, uma vez que em nenhum item este valor foi superior a 1%. Tendo em vista o anonimato dos respondentes, os questionários com múltipla marcação de respostas não são passíveis de identificação e, em consequência, impossibilitam a sua recuperação. Portanto, não nos restou outra alternativa que não fosse a eliminação dessas observações da base de dados. Este procedimento reduziu a base original de 17.709 para 16.592 casos, representando cerca de 6% de redução.

Como não foram encontrados valores extremos, passamos a analisar os valores ausentes, cuja frequência de ocorrência, para os diversos itens do questionário, está indicada na Tabela 5.

**Tabela 5.** Ocorrência de valores ausentes para os itens do questionário (N= 16.592).

<b>item</b>	<b>frequência absoluta</b>	<b>frequência relativa</b>
1	425	2,56 %
2	427	2,57 %
3	441	2,66 %
4	460	2,77 %
6	473	2,85 %
8	476	2,87 %
14	486	2,93 %
7	487	2,94 %
9	510	3,07 %
13	509	3,07 %
11	514	3,10 %
12	520	3,13 %
5	540	3,25 %
10	555	3,34 %
15	559	3,37 %
16	609	3,67 %

Os valores relatados nos revelam que o percentual de valores ausentes variou de 2,56% a 3,67%. Os resultados apresentados nos indicam, considerando cada item isoladamente, um percentual baixo de valores ausentes uma vez que em nenhum item este valor foi superior a 4%.

Resolvemos eliminar os valores ausentes pois a sua substituição pela média, regressão ou combinação de padrões, é desaconselhada nos estudos sobre modelagem por equações estruturais (Arbuckle & Wothke, 1999; Byrne, 2001). Na eliminação usamos o método *listwise*, uma vez que a eliminação dos valores ausentes pelo método *pairwise* pode impedir o alcance de uma solução convergente (Byrne, 2001). Obtivemos 1.641 casos a serem eliminados. Este número representa cerca de 9,8% do total de 16.592 observações,

pouco abaixo do limite de 10% da regra empírica de Kline (1988). A aplicação do método de eliminação listwise reduziu, então, a base de dados de 16.592 para 14.951 observações, o que ainda caracteriza uma amostra grande para os propósitos de utilização da análise fatorial e da modelagem por equações estruturais.

Para a realização das análises principais empregando análise fatorial e modelagem por equações estruturais verificamos, inicialmente, a aderência dos dados a uma distribuição normal. Para este exame recorreremos aos resultados ilustrados na Tabela 6.

**Tabela 6.** Assimetria dos itens do questionário  
(N=14.951).

item	assimetria	e.p. assimetria	razão
1	-1,67	0,02	83,50
2	-1,45	0,02	72,50
3	-1,57	0,02	78,50
4	-1,78	0,02	89,00
5	-1,92	0,02	96,00
6	-1,57	0,02	8,50
7	-1,60	0,02	80,00
8	-1,23	0,02	61,50
9	-1,48	0,02	74,00
10	-1,52	0,02	76,00
11	-1,50	0,02	75,00
12	-1,44	0,02	72,00
13	-1,52	0,02	76,00
14	-1,50	0,02	75,00
15	-1,62	0,02	81,00
16	-1,61	0,02	80,50

Uma análise superficial dos resultados relatados revela a não normalidade dos dados para todos os itens, uma vez que pelos valores na última coluna, o valor absoluto da assimetria é maior que o dobro do seu erro-padrão (Miles & Shevlin, 2001; Tabachnick & Fidell, 2001). Esta impressão inicial foi confirmada, utilizando-se os testes usualmente empregados para análise de normalidade univariada (Anderson-Darling, Ryan-Jones, Shapiro-Wilkins, Kolmogorov-Smirnov), cujos resultados indicam, do ponto de vista da significância estatística, que os dados não são aderentes a uma distribuição normal.

Uma vez que as variáveis, isoladamente, não possuem distribuição normal, elas, em conjunto, não possuem distribuição normal multivariada (Anderson, 2003). Além disso, como o valor absoluto da assimetria está entre 1 e 2, a assimetria pode ter efeito sobre a

estimação dos parâmetros (Miles & Shevlin, 2001). Portanto, nas análises subsequentes, em princípio, não empregamos métodos que exigem a normalidade dos dados.

## 5.2 Análise Preliminar

A análise preliminar da estrutura fatorial dos dados foi realizada com o auxílio do programa SPSS 13.0, empregando-se a técnica de análise fatorial. Na Tabela 7 estão indicados os valores da média(M) e desvio-padrão(DP) de cada um dos itens do questionário.

**Tabela 7.** Média e desvio-padrão dos itens do questionário (N=14.951).

<b>item</b>	<b>M</b>	<b>DP</b>
1	8,47	1,82
2	8,34	1,79
3	8,38	1,86
4	8,59	1,78
5	8,63	1,82
6	8,35	1,88
7	8,38	1,95
8	7,78	2,23
9	8,22	2,02
10	8,24	2,02
11	8,17	2,03
12	8,08	2,11
13	8,20	2,05
14	8,12	2,13
15	8,28	2,09
16	8,28	2,11

A análise dos dados revelou proximidade entre os diversos itens no tocante à cada uma das estatísticas apresentadas: média e desvio-padrão. A proximidade dos valores médios representa uma indicação inicial de que a estrutura fatorial dos dados pode ser atribuída a uma base substantiva, conceitual, e não a razões puramente estatísticas (Bernstein, Garbin & Teng, 1988). Assim, prosseguimos na análise preliminar dos dados.

A Tabela 8 apresenta a matriz de correlações entre os itens do questionário. Os resultados relatados nos mostram que os valores fora da diagonal principal são altos, indicando boa correlação entre os itens e sugerindo que, possivelmente, não teremos itens com desempenho pobre na estrutura fatorial.

**Tabela 8.** Correlações entre os itens do questionário  
(N=14.951).

item	item															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.00	0.82	0.81	0.82	0.70	0.71	0.75	0.61	0.66	0.74	0.66	0.65	0.74	0.63	0.63	0.63
2	0.82	1.00	0.85	0.82	0.71	0.74	0.77	0.69	0.71	0.75	0.70	0.70	0.78	0.68	0.67	0.67
3	0.81	0.85	1.00	0.86	0.73	0.77	0.82	0.69	0.73	0.77	0.72	0.72	0.78	0.69	0.69	0.69
4	0.82	0.82	0.86	1.00	0.77	0.76	0.80	0.65	0.69	0.75	0.69	0.68	0.76	0.67	0.67	0.66
5	0.70	0.71	0.73	0.77	1.00	0.73	0.70	0.60	0.66	0.69	0.67	0.64	0.67	0.64	0.66	0.64
6	0.71	0.74	0.77	0.76	0.73	1.00	0.76	0.68	0.72	0.76	0.71	0.71	0.73	0.68	0.68	0.70
7	0.75	0.77	0.82	0.80	0.70	0.76	1.00	0.72	0.75	0.77	0.72	0.73	0.77	0.69	0.69	0.70
8	0.61	0.69	0.69	0.65	0.60	0.68	0.72	1.00	0.76	0.75	0.76	0.75	0.73	0.74	0.70	0.71
9	0.66	0.71	0.73	0.69	0.66	0.72	0.75	0.76	1.00	0.86	0.79	0.80	0.76	0.73	0.76	0.75
10	0.74	0.75	0.77	0.75	0.69	0.76	0.77	0.75	0.86	1.00	0.81	0.81	0.80	0.75	0.76	0.77
11	0.66	0.70	0.72	0.69	0.67	0.71	0.72	0.76	0.79	0.81	1.00	0.80	0.76	0.74	0.76	0.75
12	0.65	0.70	0.72	0.68	0.64	0.71	0.73	0.75	0.80	0.81	0.80	1.00	0.78	0.76	0.77	0.81
13	0.74	0.78	0.78	0.76	0.67	0.73	0.77	0.73	0.76	0.80	0.76	0.78	1.00	0.77	0.73	0.73
14	0.63	0.68	0.69	0.67	0.64	0.68	0.69	0.74	0.73	0.75	0.74	0.76	0.77	1.00	0.80	0.74
15	0.63	0.67	0.69	0.67	0.66	0.68	0.69	0.70	0.76	0.76	0.76	0.77	0.73	0.80	1.00	0.78
16	0.63	0.67	0.69	0.66	0.64	0.70	0.70	0.71	0.75	0.77	0.75	0.81	0.73	0.74	0.78	1.00

sign

determinante da matriz: 0, 00000000424

\* valores são todos nulos

Os resultados relativos à parte da significância são zeros, o que reforça os indícios de fatorabilidade da matriz (Pasquali, 2005). O valor do determinante da matriz, embora extremamente baixo, não é rigorosamente zero. Portanto, temos a indicação da presença de mais de um fator, ou seja, temos a indicação de que fatorar a matriz é pertinente (Pasquali, 2005).

Na Tabela 9 temos os dados referentes ao índice de Kaiser-Meyer-Olkin (KMO) e ao teste de esfericidade de Bartlett. O valor 0,98 para o KMO, próximo de 1, pode ser considerado muito bom e representa mais um indício de fatorabilidade da matriz (Laros & Puente-Palácios, 2004). Os valores relatados nos mostram a significância do teste, o que reforça a possibilidade de fatorabilidade da matriz (Pasquali, 2005).

**Tabela 9.** Índice KMO e teste de esfericidade de Bartlett  
(N=14.951).

	KMO	0,98
teste de	qui-quadrado (aprox.)	288083,08
Bartlett	graus de liberdade	120
	significância	< 0,01

Prosseguindo na análise tivemos que decidir o número de fatores a ser extraído. Na Tabela 10 ilustramos os resultados obtidos com o critério da análise paralela. Os valores

relatados foram obtidos com o uso da rotina do SPSS disponível em O'Connor (2000), utilizando o método *principal axis factorization* (PAF) e permutação dos dados empíricos. Esta escolha deveu-se à não normalidade dos dados empíricos. Os valores mostrados na coluna dos autovalores aleatórios representam as médias dos autovalores obtidos com 10 conjuntos de dados e um intervalo de confiança de 95%, que são as alternativas padrão do software utilizado.

**Tabela 10.** Autovalores empíricos e aleatórios dos primeiros 16 componentes (N=14.951).

componente	autovalor empírico	autovalor aleatório
1	11,708	0,056
2	0,655	0,044
3	0,134	0,037
4	0,100	0,029
5	0,057	0,021
6	0,034	0,016
7	0,009	0,010
8	-0,001	0,005
9	-0,015	-0,001
10	-0,034	-0,006
11	-0,042	-0,014
12	-0,047	-0,020
13	-0,064	-0,029
14	-0,070	-0,035
15	-0,076	-0,043
16	-0,092	-0,055

Pelos resultados relatados podemos verificar que o componente 6 é o último componente em que os valores empíricos são superiores aos valores aleatórios, o que nos sugere a existência de 6 fatores. Assim, deveríamos prosseguir com a análise preliminar e extrair 6 fatores.

Antes de prosseguir vamos fazer uma pequena reflexão. Em primeiro lugar, talvez devamos considerar que pesquisa não é Estatística (Bachrach, 1969). A Estatística, assim, como a Matemática, deve ser vista como um holofote que ilumina, caso necessário, um caminho escuro, mas não deve ser utilizada como substituta do caminho. Estatística e Matemática devem ser utilizadas, primordialmente, como instrumentos úteis para o tratamento de dados (Alves & Peres-dos-Santos, 1977, 1980a, 1980b; Peres-dos-Santos, 1969, 1975, 1976a, 1976b, 1978, 1980, 1981, 1987; Peres-dos-Santos & Laros, 2007).

Em segundo lugar, sabemos que dificilmente um pesquisador coleta e analisa dados sem ter uma idéia *a priori* do relacionamento entre as variáveis (Floyd & Widaman, 1995).

Assim, um fator deve representar algo relevante no contexto da teoria científica e não apenas se posicionar como um parâmetro matemático (Pasquali, 2005). O critério da relevância do fator talvez seja o critério mais interessante e útil, pois reter um fator que tenha apenas significado matemático não faz sentido em ciência (Pasquali, 2005). Desta forma, o número de fatores a serem extraídos, indicado por algum critério de extração, deve ser calibrado pelo critério da relevância do fator. Portanto, a aplicação do critério da relevância do fator nos sugere que o número de fatores (6), sugerido por um método puramente matemático ou estatístico é totalmente inadequado e devemos descartá-lo. Afinal de contas, é melhor ficar sem resposta do que aceitar uma resposta inadequada (Skinner, citado por Bachrach, 1969).

Por outro lado, essa decisão poderia ser criticada por talvez representar a chamada ortodoxia da teimosia (Gardner, citado por Bachrach, 1969) ou miopia de hipótese (Bachrach, 1969). Nesta ortodoxia ou miopia, o pesquisador possui uma incapacidade de ver longe ao pesquisar, devido a idéias pré-concebidas. Por este distúrbio de visão, o pesquisador não mantém os olhos abertos. O pesquisador limita-se a um caminho fixo, a ponto de ficar aparentemente cego para eventos não usuais que possam vir a ocorrer. O pesquisador não prossegue na busca de novos fatos. O pesquisador diante dos fatos, recusa-se a aceitá-los ou procura descartar-se deles com explicações simplistas.

Nesta situação, com o pesquisador ocorre o fenômeno em que a teoria começa a dominar os dados e ele começa a preocupar-se em provar que sua teoria é correta por meio da investigação dos dados que podem ou não apoiá-la (Bachrach, 1969). Este é o posicionamento das chamadas pessoas ortodoxas (Renan, citado por Bachrach, 1969). Para este autor, as pessoas ortodoxas tem geralmente muita pouca honestidade científica: não investigam, tentam provar que alguma coisa tem que ser como dizem; os resultados lhe foram fornecidos de antemão e este é o certo, indubitavelmente certo.

O melhor é o pesquisador, ao encontrar alguma coisa de interessante, deixar tudo o mais para estudá-la (Skinner, citado por Bachrach, 1969). Afinal de contas, Ciência é um conjunto de atitudes, uma disposição para lidar com os fatos e não com a opinião que alguém emitiu sobre eles (Skinner, citado por Bachrach, 1969).

Aparentemente estamos diante de um dilema, uma vez que não podemos simplesmente abandonar o número obtido, pois em ciência, sagrados e intocáveis são os fatos e não as teorias (Pasquali, 2005), e o assunto estudado sabe mais que os homens (Skinner, citado por Bachrach, 1969), ou seja, os dados subsistem, não os homens (Bachrach, 1969). Os

experimentos nem sempre dão os resultados que se espera, mas os fatos devem permanecer e as expectativas desvanecerem-se (Skinner, citado por Bachrach, 1969). Devemos rejeitar explicações prematuras e esperar por uma resposta combinada com uma busca contínua de explicação, pois todo dado obtido fornece informação para a mente preparada que respeita os dados e não deixa que as hipóteses impeçam a pesquisa (Bachrach, 1969). O que fazer?

Nosso posicionamento é o de que um estudo científico deve representar um equilíbrio entre a observação e o experimento, uma vez que a primeira fornece a coleta empírica dos dados e o segundo representa o raciocinar sobre estes dados visando obter maiores conhecimentos. Afinal de contas, a razão se amplia na experimentação mas está enraizada na observação (Bachrach, 1969). Além disso, nenhum método analítico, inclusive análise fatorial, nos dita como nossos construtos ou teorias deveriam ser (Mulaik, citado por Laros, 2005). Portanto, como somos nós que criamos significados para as coisas, decidimos utilizar o critério de relevância para calibrar o número de fatores a serem extraídos.

Como o modelo desenvolvido nesta tese postula a existência de dois fatores e estamos, ainda, numa fase de análise preliminar, vamos considerar os resultados provenientes da extração de 1, 2 e 3 fatores, seguindo a sugestão de Osborne e Costello (2005). A Tabela 11 ilustra os resultados obtidos com o SPSS, usando-se o método *principal axis factoring* (PAF).

**Tabela 11.** Fatores e percentual de variância explicada (N=14.951).

fatores	autovalor	variância explicada (%)	variância explicada acumulada (%)
1	11,7	73,0	73,0
1	11,7	73,2	
2	0,7	4,2	77,4
1	11,7	73,3	
2	0,7	4,3	
3	0,2	1,0	78,6

A Tabela 11 nos mostra que em todas as extrações o primeiro fator é responsável por mais de 70% de explicação da variância, o que é bastante satisfatório (Gorsuch, 1983). O ganho em passar da solução de 1 para 2 fatores é de 4,4% em termos de explicação da variância às custas de se ficar com um autovalor menor que 1. Na passagem de 1 para 3 fatores ganha-se 5,6% na explicação da variância e ficam-se com dois autovalores menores que 1. Na passagem de 2 para 3 fatores ganha-se 1,20% na explicação da variância porém

fica-se com mais um autovalor menor que 1. Os resultados relatados mostram, ainda, que o terceiro autovalor obtido é muito inferior a 1. Na solução com três fatores, o último não atende ao critério da importância do fator, que exige um mínimo de 3% de explicação da variância (Harman, 1976). Assim, a solução com dois fatores, além de ser bastante satisfatória, nos parece a mais promissora. Esta solução foi a adotada na continuação da análise preliminar.

Uma vez definidos o número de fatores e o método de extração, tivemos que escolher o método de rotação para se obter uma solução fatorial. Como podemos defender o argumento teórico de que os fatores estão relacionados, o caminho natural foi escolher um método de rotação oblíqua em busca de uma solução satisfatória. Entre os diversos métodos disponíveis para a rotação oblíqua não há um método que seja majoritariamente utilizado em detrimento dos demais (Costello & Osborne, 2005). No caso particular do SPSS pareceu ser uma boa diretriz manter os valores padrão disponíveis nos diversos métodos, uma vez que não há relatos de ganhos pela utilização de valores diferentes (Fabrigar & cols., 1999).

Desta forma, utilizamos o método de extração PAF e rotação Promax com as opções padrão do software (Promax, Kaiser normalization, kappa = 4). Obtivemos uma solução convergente após 3 iterações. Os resultados obtidos encontram-se descritos na Tabela 12, em que estão omitidas as cargas fatoriais cujos valores absolutos são inferiores a 0,32 (Laros, 2005; Laros & Puente-Palacios, 2004; Pasquali, 2005, 2006; Tabachnick & Fidell, 2001).

**Tabela 12.** Cargas fatoriais e comunalidade dos itens do questionário (N=14.951).

item	descrição do item	Fator		
		F1	F2	$h^2$
1	responde adequadamente às perguntas dos alunos	0,90	-	78%
2	promove a interação entre os alunos	0,82	-	80%
3	procura criar uma atmosfera agradável na sala de aula	0,84	-	86%
4	mantém relacionamento cordial com os alunos	0,94	-	86%
5	aplica as mesmas regras para todos os alunos	0,63	-	65%
6	resolve, com habilidade, os problemas de indisciplina em sala de aula.	0,55	0,34	72%
7	ministra as aulas com entusiasmo e bom humor	0,61	-	75%
8	propõe tarefas interessantes para favorecer o aprendizado	-	0,75	70%
9	usa uma linguagem clara na apresentação da matéria	-	0,77	79%
10	responde aos questionamentos dos alunos de forma clara	-	0,68	83%
11	orienta, com clareza, os trabalhos propostos	-	0,79	78%
12	desenvolve a matéria num ritmo satisfatório	-	0,90	82%
13	encoraja os alunos a fazer questionamentos	0,40	0,53	78%
14	propõe provas que valorizam a reflexão mais do que a memorização	-	0,80	74%
15	demonstra coerência entre as provas e os conteúdos desenvolvidos em sala de aula	-	0,83	75%
16	utiliza o tempo de aula de forma produtiva	-	0,83	75%
Variância explicada com os dois fatores				80,1%
Correlação entre os fatores				0,81

A análise da Tabela 12 nos indica que, com exceção dos itens 6 e 13, os demais itens são todos itens simples, isto é, possuem cargas fatoriais principais em apenas um dos fatores. Sabemos, no entanto, que a interpretabilidade da estrutura fatorial é otimizada quando a estrutura é simples (Thurstone, citado por Pasquali, 2005), os itens não possuem cargas similares em dois ou mais fatores (Gorsuch, 1983) e um item deve ser excluído quando as diferenças entre os valores absolutos das cargas fatoriais principais é menor que 0,10 (Laros & Puentes-Palácio, 2004). A aplicação destes critérios nos leva a manter o item

6 no fator 1 e o item 13 no fator 2. Esta decisão é apoiada também no fato de que tão importante quanto investigar as características psicométricas do item é investigar se ele consegue representar, de maneira acurada, o conteúdo abordado no fator (Smith & McCarthy, 1995).

Concluimos, portanto, pelo agrupamento dos itens 1 a 7 num primeiro fator, com cargas fatoriais variando de 0,55 a 0,94 e o agrupamento dos itens 8 a 16 num outro fator, com cargas variando de 0,53 a 0,90. Todos os fatores podem ser considerados sólidos, uma vez que atendem ao requisito de possuírem no mínimo cinco itens com cargas fatoriais mínimas de 0,50 (Costello & Osborne, 2005).

Observamos, ainda, uma alta correlação entre os dois fatores: 0,81. Essa alta correlação sugere, possivelmente, a existência de um fator geral de segunda ordem. Na Tabela 13 apresentamos os resultados obtidos com a extração de um único fator.

Os resultados relatados nos mostram que as cargas fatoriais variam de 0,79 a 0,90, demonstrando a solidez do fator (Costello & Osborne, 2005). Há uma predominância, em termos de cargas fatoriais, pelos itens que constituem o fator 2, anteriormente descrito, uma vez que, dos seis itens com maiores cargas fatoriais, quatro deles, representando cerca de 67% do total, são itens correspondentes ao fator 2. Os valores das correlações item-total, na faixa de 0,81 a 0,89, excetuando-se o item cinco, cujo valor é 0,78, são bons (Hair & cols., 2005). Os dezesseis itens, representativos do fator, constituem uma escala cuja consistência interna, medida pelo alfa de Cronbach, é 0,96, o que representa um valor muito bom (Nunnally & Berstein, 1994; Pasquali, 2005).

**Tabela 13.** Cargas fatoriais, comunalidade e correlação item-total para os itens do fator de segunda ordem (N=14.951).

item	descrição do item O professor:	Fator		
		F1	$h^2$	$r_{it}$
10	responde aos questionamentos dos alunos de forma clara	0,90	0,82	0,89
3	procura criar uma atmosfera agradável na sala de aula	0,89	0,78	0,87
13	encoraja os alunos a fazer questionamentos	0,88	0,78	0,87
7	ministra as aulas com entusiasmo e bom humor	0,87	0,76	0,86
9	usa uma linguagem clara na apresentação da matéria	0,87	0,76	0,86
12	desenvolve a matéria num ritmo satisfatório	0,87	0,76	0,86
2	promove a interação entre os alunos	0,86	0,74	0,85
11	orienta, com clareza, os trabalhos propostos	0,86	0,74	0,85
4	mantém relacionamento cordial com os alunos	0,86	0,74	0,85
6	resolve os problemas de indisciplina em sala de aula	0,85	0,72	0,83
15	demonstra coerência entre as provas e os conteúdos desenvolvidos em sala de aula	0,84	0,70	0,83
16	utiliza o tempo de aula de forma produtiva	0,84	0,70	0,83
14	propõe provas que valorizam a reflexão mais do que a memorização	0,83	0,70	0,83
1	responde adequadamente às perguntas dos alunos	0,82	0,68	0,81
8	propõe tarefas interessantes para favorecer o aprendizado	0,82	0,67	0,81
5	aplica as mesmas regras para todos os alunos	0,79	0,63	0,78
Variância explicada		73,0%		
Coeficiente de fidedignidade		0,98		

Retornando aos resultados apresentados na Tabela 12 e utilizando os valores das cargas fatoriais obtidas podemos estabelecer uma hierarquia, em relação à contribuição dos diversos itens para o respectivo fator. Esta hierarquia acha-se descrita na Tabela 14 para o Fator 1: relacionamento professor-aluno.

**Tabela 14.** Cargas fatoriais dos itens que fazem parte do fator 1 (N=14.951).

descrição do item		
item	O professor:	carga
4	mantém relacionamento cordial com os alunos	0,94
1	responde adequadamente às perguntas dos alunos	0,90
3	procura criar uma atmosfera agradável na sala de aula	0,84
2	promove a interação entre os alunos	0,82
5	aplica as mesmas regras para todos os alunos	0,63
7	ministra as aulas com entusiasmo e bom humor	0,61
6	resolve os problemas de indisciplina em sala de aula	0,55

Os resultados relatados mostram que os itens mais importantes na interpretação do primeiro fator foram os relativos ao relacionamento cordial do professor com os alunos e a resposta adequada às perguntas dos alunos. O item com menos importância na interpretação do fator foi o item referente à resolução dos problemas de indisciplina em sala de aula. Os resultados obtidos indicam que 57% das cargas são excelentes, 14% são muito boas e 29% são boas (Comrey, citado por Pasquali, 2005).

Na Tabela 15 mostramos a hierarquia das cargas fatoriais para o Fator 2: prática de ensino. Os resultados relatados revelam que os itens mais importantes na interpretação do segundo fator foram os relativos ao desenvolvimento da matéria num ritmo satisfatório, demonstração de coerência entre as provas e os conteúdos desenvolvidos em sala de aula e a utilização do tempo de aula de forma produtiva. O item com menos importância na interpretação do fator foi o encorajamento dos alunos a fazer questionamentos. Quanto às cargas fatoriais obtidas, temos que 78% delas são excelentes, 11% são muito boas e 11% são razoáveis (Comrey, citado por Pasquali, 2005).

**Tabela 15.** Cargas fatoriais dos itens que fazem parte do fator 2 (N=14.951).

descrição do item		
item	O professor:	carga
12	desenvolve a matéria num ritmo satisfatório	0,90
15	demonstra coerência entre as provas e os conteúdos desenvolvidos em sala de aula	0,83
16	utiliza o tempo de aula de forma produtiva	0,83
14	propõe provas que valorizam a reflexão mais do que a memorização	0,80
11	orienta, com clareza, os trabalhos propostos	0,79
9	usa uma linguagem clara na apresentação da matéria	0,77
8	propõe tarefas interessantes para favorecer o aprendizado	0,75
10	responde aos questionamentos dos alunos de forma clara	0,68
13	encoraja os alunos a fazer questionamentos	0,53

Algumas características psicométricas do fator 1 são mostradas na Tabela 16.

**Tabela 16.** Média, desvio-padrão e correlação item-total para os itens do fator 1 (N=14.951).

descrição do item				
item	O professor:	M	DP	$r_{it}$
5	aplica as mesmas regras para todos os alunos	8,63	1,82	0,79
4	mantém relacionamento cordial com os alunos	8,59	1,78	0,90
1	responde adequadamente às perguntas dos alunos	8,47	1,82	0,85
3	procura criar uma atmosfera agradável na sala de aula	8,38	1,86	0,90
7	ministra as aulas com entusiasmo e bom humor	8,38	1,95	0,85
6	resolve os problemas de indisciplina em sala de aula	8,35	1,88	0,83
2	promove a interação entre os alunos	8,34	1,79	0,88
Coeficiente de fidedignidade				0,96

A análise dos resultados mostra que o intervalo de variação das médias dos itens é o intervalo de 8,34 a 8,63, com a variação do desvio-padrão representada pela faixa de 1,78 a 1,95. Estes resultados indicam que os professores receberam avaliações muito favoráveis no fator relacionamento professor-aluno. Os itens mais positivamente avaliados pelos alunos, com as maiores médias, foram os relativos à manutenção de um relacionamento cordial com os alunos e aplicação da mesma regra para todos os alunos. O item menos positivamente avaliado, apesar de uma boa média, foi o relativo à resolução, por parte do professor, dos problemas de indisciplina em sala de aula. O item que apresentou a maior variabilidade absoluta é “ministra as aulas com entusiasmo e bom humor”.

Os valores da correlação item-total, na faixa de 0,83 a 0,90 são bons (Hair & cols., 2005). Os sete itens, representativos do fator, constituem uma escala cuja consistência interna, medida pelo alfa de Cronbach, é 0,96, o que representa um valor muito bom (Nunnally & Bernstein, 1994; Pasquali, 2005).

Algumas características psicométricas do fator 2 são mostradas na Tabela 17.

**Tabela 17.** Média, desvio-padrão e correlação item-total para os itens do fator 2 (N=14.951).

descrição do item		M	DP	$r_{it}$
item	O professor:			
15	demonstra coerência entre as provas e os conteúdos desenvolvidos em sala de aula	8,28	2,09	0,85
16	utiliza o tempo de aula de forma produtiva	8,28	2,11	0,85
10	responde aos questionamentos dos alunos de forma clara	8,24	2,02	0,89
9	usa uma linguagem clara na apresentação da matéria	8,22	2,02	0,87
13	encoraja os alunos a fazer questionamentos	8,20	2,05	0,85
11	orienta, com clareza, os trabalhos propostos	8,17	2,03	0,87
14	propõe provas que valorizam a reflexão mais do que a memorização	8,12	2,13	0,84
12	desenvolve a matéria num ritmo satisfatório	8,08	2,11	0,89
8	propõe tarefas interessantes para favorecer o aprendizado	7,78	2,23	0,82
Coeficiente de fidedignidade				0,97

A análise dos resultados mostra que o intervalo de variação das médias dos itens é o intervalo de 7,78 a 8,28, com a variação do desvio-padrão representada pela faixa de 2,02 a 2,23. Estes resultados indicam que os professores receberam avaliações favoráveis no fator prática de ensino. Os itens mais positivamente avaliados pelos alunos, com as maiores médias, foram os relativos à demonstração de coerência entre as provas e os conteúdos desenvolvidos em sala de aula e à utilização do tempo de aula de forma produtiva. O item menos positivamente avaliado foi o relativo à proposição de tarefas interessantes para favorecer o aprendizado. O item que apresentou a maior variabilidade absoluta foi o relativo à proposição de tarefas interessantes para favorecer o aprendizado.

Os valores da correlação item-total, na faixa de 0,82 a 0,89 são bons (Hair & cols., 2005). Os resultados mostram, ainda, que os 9 itens, representativos do fator, constituem uma escala cuja consistência interna, medida pelo alfa de Cronbach, é 0,97, o que representa um valor muito bom (Nunnally & Bernstein, 1994; Pasquali, 2005).

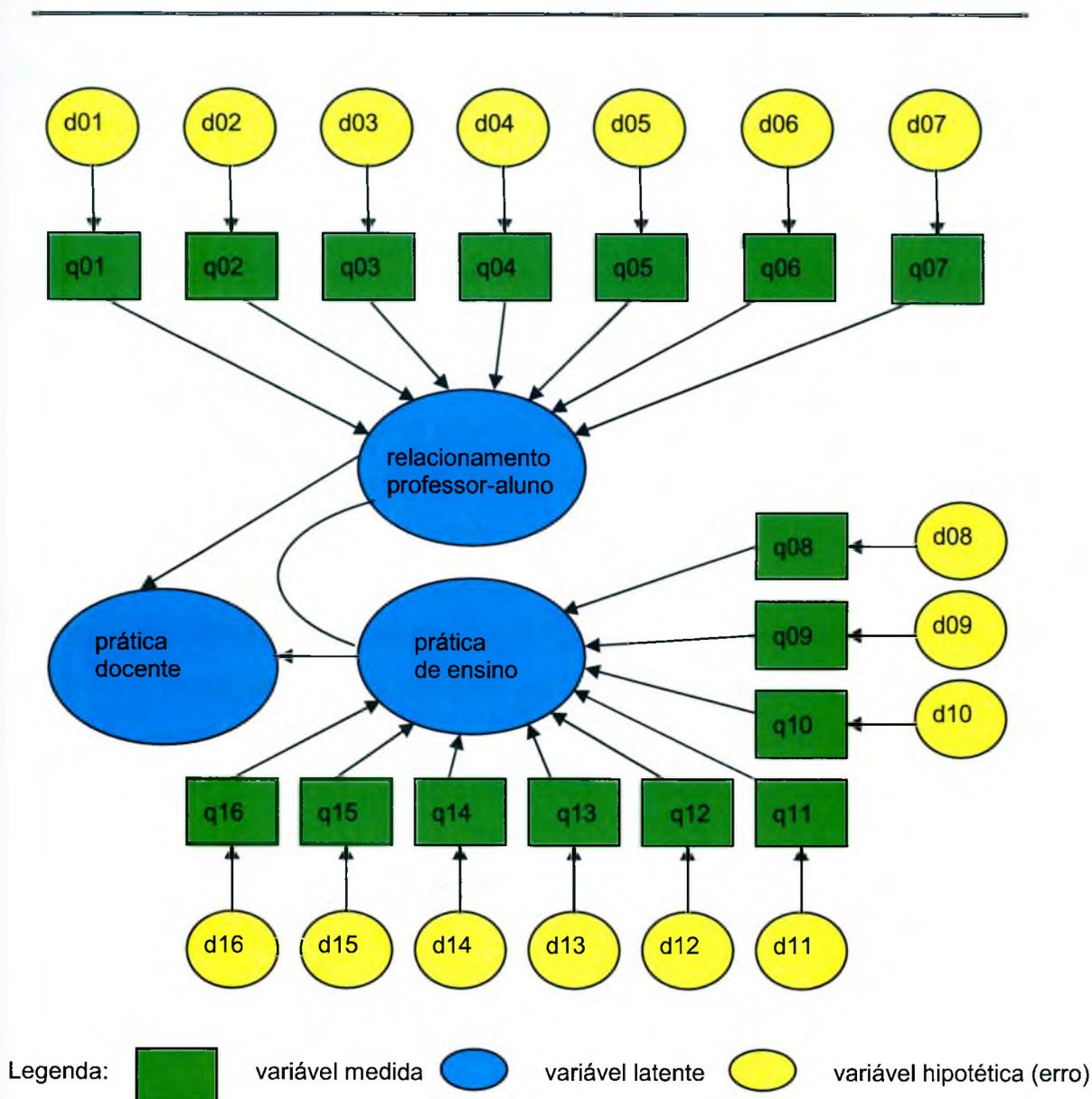
Os resultados das Tabelas 16 e 17 sugerem ainda que os professores receberam avaliações mais favoráveis nos itens referentes ao fator relacionamento professor-aluno do que nos itens referentes ao fator prática de ensino.

### 5.3 Análise Confirmatória

A análise confirmatória da estrutura fatorial dos dados foi realizada com o auxílio do programa AMOS 4.0, empregando-se a técnica de modelagem por equações estruturais. O modelo objeto desta análise, acha-se ilustrado na Figura 6 e foi baseado no modelo postulado na Figura 2, corroborado pela análise fatorial realizada na fase preliminar.

Na Figura 6, as variáveis relacionamento professor-aluno, prática de ensino e prática docente, representadas por elipses, são latentes. As variáveis q01, q02, ..., q16, correspondentes às variáveis descritas nos itens 1 a 16 do questionário, representadas por retângulos, são variáveis medidas (observadas). As variáveis d1, d2, ..., d16, representadas por elipses, representam variáveis hipotéticas, destinadas a absorver a variação aleatória inerente a todo processo que utiliza dados empíricos e são representativas dos erros de medida e de variáveis que não foram contempladas no modelo. As relações explicativas entre as diversas variáveis do modelo são representadas por setas unidirecionais. Das três variáveis latentes, a variável prática docente, por não receber setas, é uma variável exógena, enquanto que as variáveis relacionamento professor-aluno e prática de ensino são

endógenas (Pasquali, 2006). O relacionamento entre os fatores relacionamento professor-aluno e prática de ensino é representado pelo arco entre os dois fatores.



**Figura 6.** Modelo estrutural de avaliação do professor pelo aluno (modelo teórico).

O modelo indicado na Figura 6 compreende dois submodelos. O primeiro engloba as variáveis q01, q02, q03, q04, q05, q06, q07 e representa o modelo de mensuração correspondente à variável latente relacionamento professor-aluno. O segundo modelo engloba as variáveis q08, q09, q10, q11, q12, q13, q14, q15, q16 e representa o modelo de mensuração correspondente à variável latente prática de ensino.

Após a especificação do modelo, temos que analisar a sua possibilidade de identificação, antes da estimação dos parâmetros (Boomsma, 2000; McDonald & Ho, 2002). Uma conhecida, forte e suficiente condição para que os parâmetros do modelo sejam identificados, exceto quanto a aspectos de escala, é a condição de que as cargas fatoriais formem agrupamentos independentes (McDonald & Ho, 2002). Num agrupamento independente cada variável observada possui carga fatorial elevada em apenas um fator. Isto exige que cada variável latente ou fator comum tenha pelo menos dois indicadores puros se os fatores são correlacionados ou pelos menos três se eles são correlacionados (McDonald & Ho, 2002). Assim, numa base substantiva, corroborada pela análise fatorial preliminar, estas condições são atendidas pelos modelos de mensuração indicados na Figura 6.

Em relação à escala decidimos pela utilização de escalas padronizadas. Esta decisão é apoiada pela prática da literatura da área que mostra que a utilização de uma solução completamente padronizada, em termos de escala, também ajuda na interpretação dos resultados obtidos (McDonald & Ho, 2002).

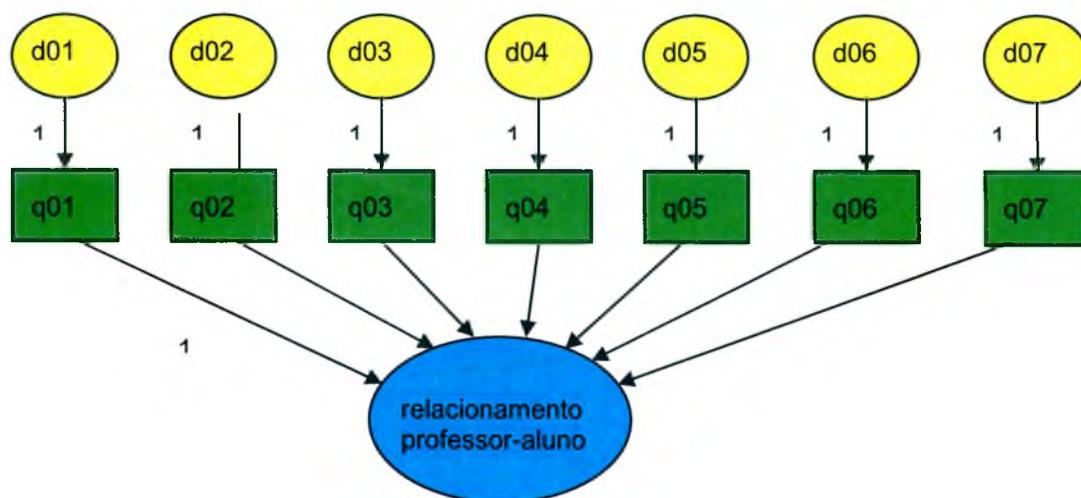
Após analisadas a especificação e possibilidade de identificação do modelo, temos que analisar os pressupostos da modelagem por equações estruturais relativos a amostra, valores ausentes e método de estimação. Os aspectos de amostra grande e valores ausentes já foram tratados na análise exploratória dos dados. A decisão tomada em seguida referiu-se à escolha do método de estimação dos parâmetros, dentre os principais disponíveis no programa AMOS 4.0: *asymptotically distribution-free* (ADF), *generalized least squares* (GLS) e *maximum likelihood* (ML).

Cada um desses métodos tem suas especificidades, vantagens e desvantagens (Byrne, 2001). O método de livre distribuição assintótica (ADF) não faz exigências quanto à natureza da distribuição envolvida, enquanto os métodos de máxima verossimilhança (ML) e mínimos quadrados generalizados (GLS), que são os métodos empregados com maior frequência na modelagem por equações estruturais, exigem que os dados sejam contínuos e tenham distribuição normal multivariada (Bollen, 1989; Byrne, 2001). O método ADF exige mais recursos computacionais que os outros dois métodos (Byrne, 2001). Como vimos anteriormente, os nossos dados não atendem aos pressupostos de normalidade o que, em princípio, inviabiliza o utilização dos métodos ML ou GLS. Utilizando, no entanto, o procedimento de reamostragem (Efron, 1979; Efron & Tibshirani, 1993), podemos contornar esta limitação (Byrne, 2001). Utilizamos, então, na estimação dos parâmetros, o

procedimento de reamostragem, com 500 amostras, usando o método ML, conforme sugestão de Byrne (2001).

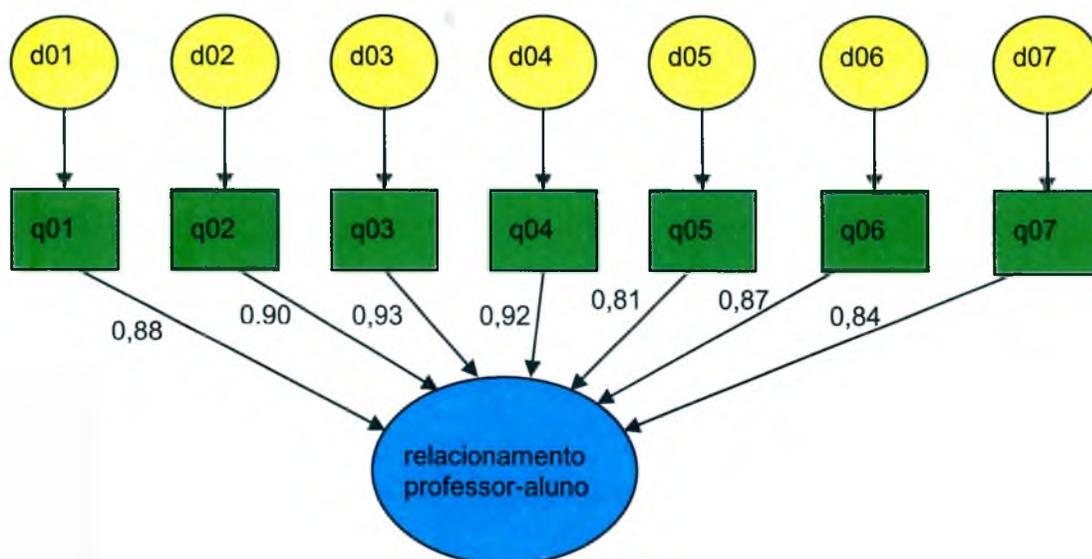
Na análise propriamente dita do modelo indicado na Figura 6, seguimos a sugestão de Kline (1998) e Pasquali (2006). Para estes autores, a análise deve ser feita por partes, verificando-se, inicialmente, de forma *ad hoc*, a convergência e consistência de cada modelo de mensuração. Em seguida devem ser analisados os modelos estruturais parciais e, finalmente, o modelo estrutural completo.

Assim, a primeira análise realizada foi a do modelo de mensuração relativa ao relacionamento professor-aluno, indicado na Figura 7. Os valores 1 indicados em algumas setas no diagrama são necessários para que o modelo proposto seja identificável, do ponto de vista computacional.



**Figura 7.** Modelo de mensuração relativo ao fator relacionamento professor-aluno (modelo teórico).

A análise do modelo, utilizando-se o programa AMOS 4.0, método de estimação ML e procedimento de reamostragem, com 500 amostras, forneceu os resultados indicados na Figura 8, em que as cargas fatoriais aparecem numa escala padronizada.



**Figura 8.** Modelo de mensuração relativo ao fator relacionamento professor-aluno.

Os resultados da análise deste modelo estão na Tabela 18.

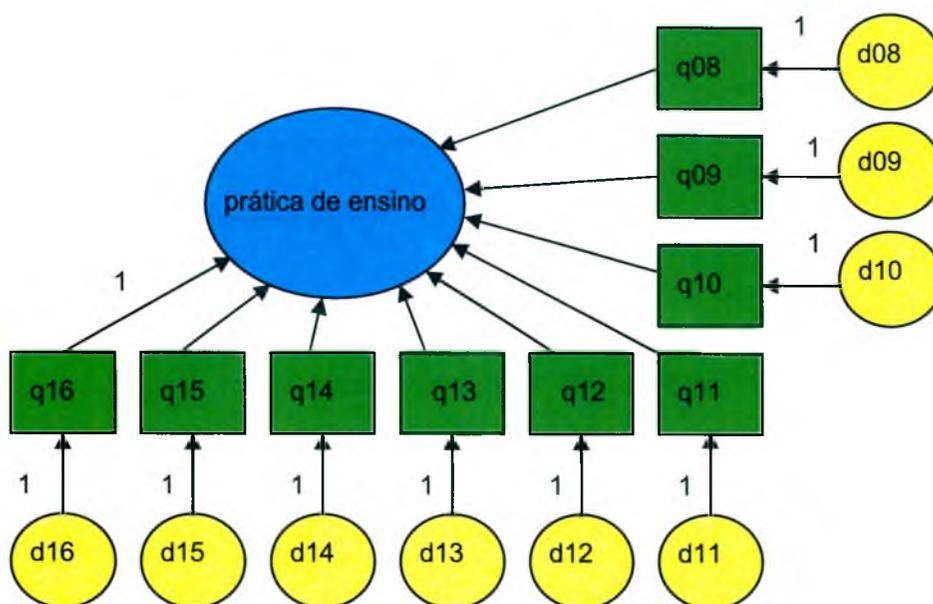
**Tabela 18.** Análise confirmatória do modelo de mensuração relativo ao relacionamento professor-aluno (N=14.951)

variável	carga fatorial		
q03	0,93		
q04	0,92		
q02	0,90		
q01	0,88		
q06	0,87		
q07	0,84		
q05	0,81		
índices de adequação			
CFI	0,981	intervalo de confiança (0,90)	
GFI	0,961	limite inferior	limite superior
RMSEA	0,100	0,096	0,104

Os resultados apresentados na Tabela 18 revelam que cinco cargas, representando cerca de 71,4% do total, estão acima de 0,85 e as outras duas cargas, representando 28,6% do total, estão entre 0,81 e 0,84. Assim, do ponto de vista das cargas fatoriais, o modelo pode ser considerado como bom. Quanto aos índices de adequação, o valor do índice

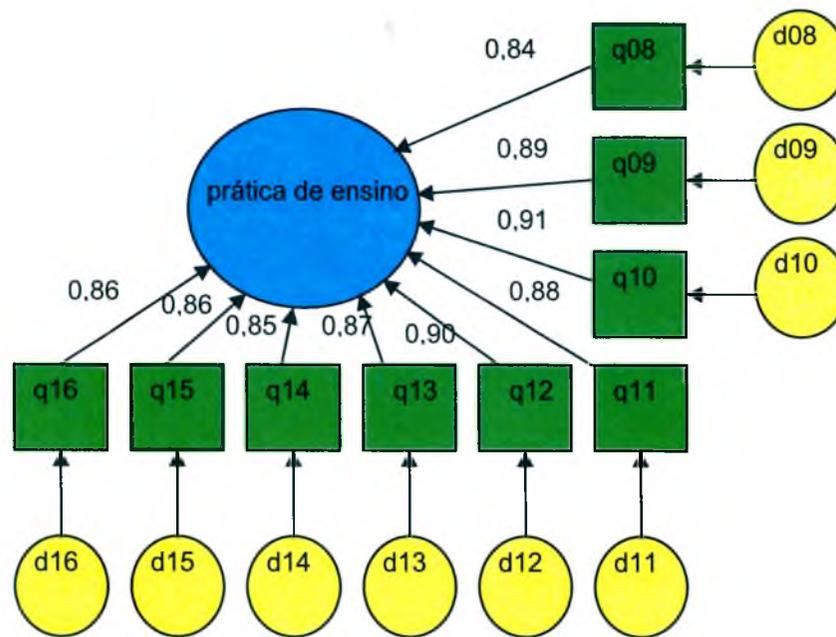
RMSEA, no limiar do limite de rejeição, indica uma adequação pobre do modelo. Os demais índices, CFI e GFI, por estarem próximos de 1, podem ser considerados como bons. Podemos, portanto, considerar os índices, no seu conjunto, como satisfatórios. Como não há razão substantiva para a sua reespecificação, podemos considerar o modelo de mensuração relativo ao fator relacionamento professor-aluno como satisfatório.

A segunda análise realizada foi a do modelo de mensuração relativa à prática de ensino, indicado na Figura 9. Os valores 1 indicados em algumas setas no diagrama são necessários para que o modelo proposto seja identificável, do ponto de vista computacional.



**Figura 9.** Modelo de mensuração relativo ao fator prática de ensino (modelo teórico).

A análise do modelo, utilizando-se o programa AMOS 4.0, método de estimação ML e procedimento de reamostragem, com 500 amostras, forneceu os resultados indicados na Figura 10, em que as cargas fatoriais aparecem numa escala padronizada.



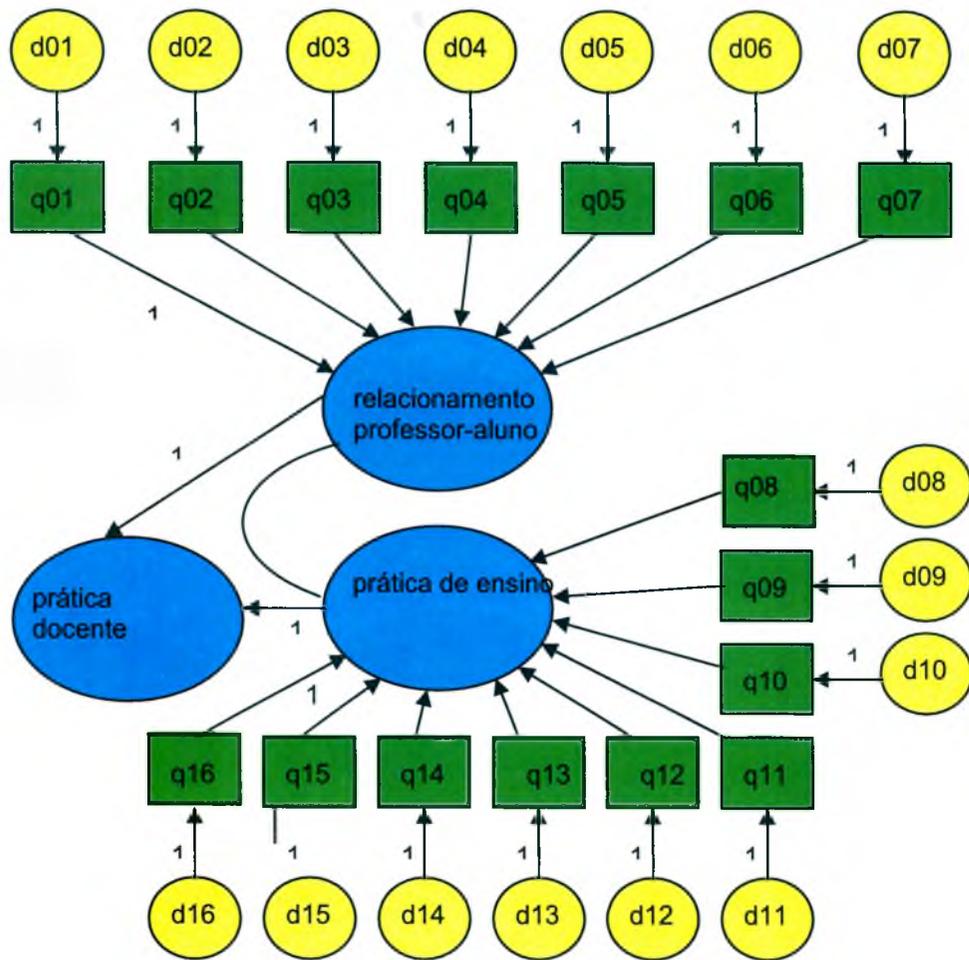
**Figura 10.** Modelo de mensuração relativo ao fator prática de ensino.

Os resultados da análise deste modelo estão na Tabela 19. Os resultados relatados revelam que sete cargas, representando cerca de 77,8% do total, estão acima de 0,85 e as outras duas cargas, representando 22,2% do total, estão entre 0,84 e 0,85. Assim, do ponto de vista das cargas fatoriais, o modelo pode ser considerado como bom. Quanto aos índices de adequação, o valor do índice RMSEA, no limiar do limite de rejeição, indica uma adequação pobre do modelo. Os demais índices, CFI e GFI, por estarem próximos de 1, podem ser considerados como bons. Podemos, portanto, considerar os índices, no seu conjunto, como satisfatórios. Como não há razão substantiva para a sua reespecificação, podemos considerar o modelo de mensuração relativo ao fator prática de ensino como satisfatório.

**Tabela 19.** Análise confirmatória do modelo de mensuração relativo à prática de ensino (N=14.951).

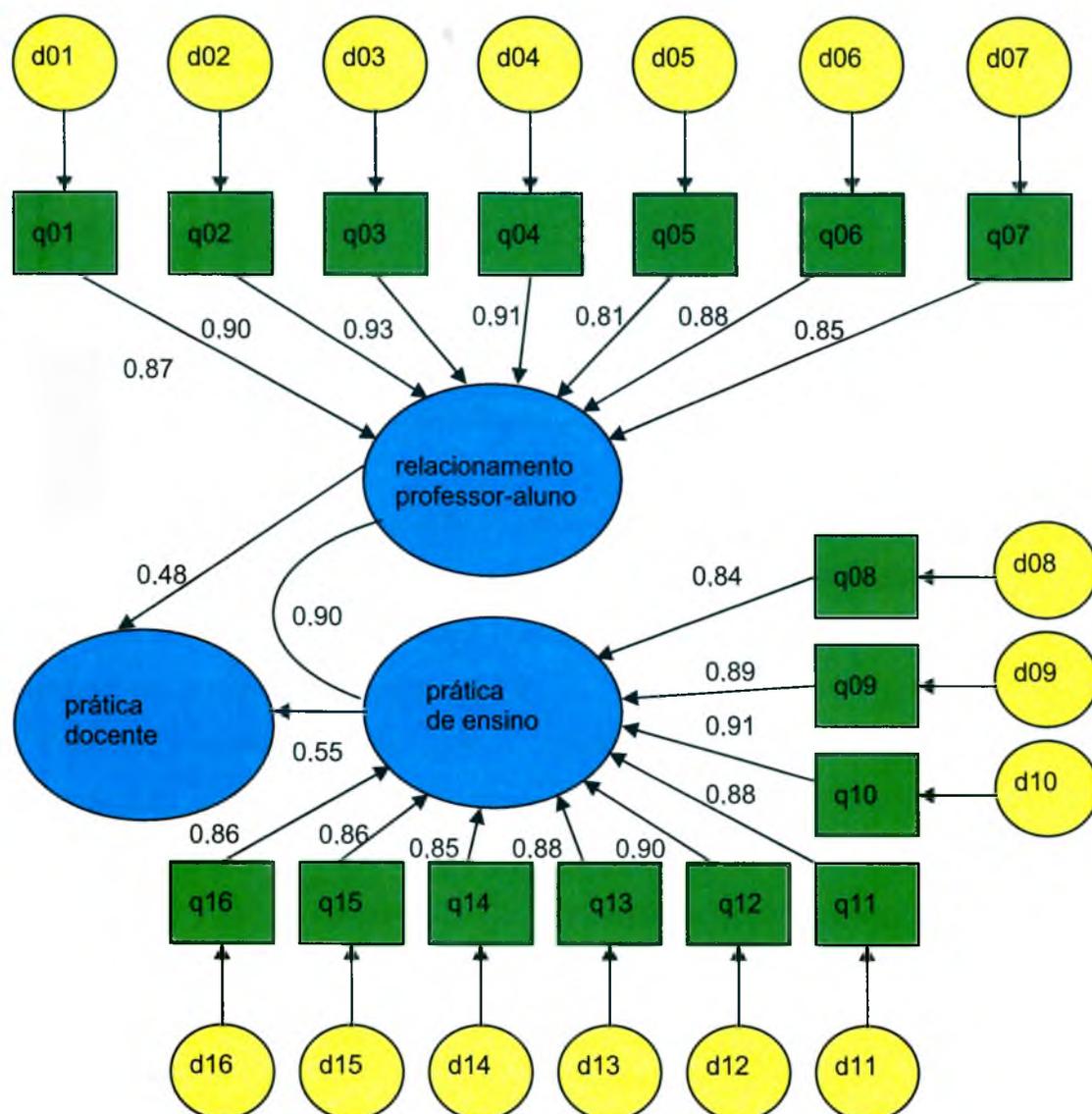
variável	carga fatorial		
q10	0,91		
q12	0,90		
q09	0,89		
q11	0,88		
q13	0,87		
q15	0,86		
q16	0,86		
q14	0,85		
q08	0,84		
índices de adequação		intervalo de confiança (0,90)	
CFI	0,970	limite inferior	limite superior
GFI	0,936		
RMSEA	0,106	0,104	0,109

Finalmente procedemos à análise do modelo estrutural completo, indicado na Figura 11. Os valores 1 indicados em algumas setas no diagrama são necessários para que o modelo proposto seja identificável, do ponto de vista computacional.



**Figura 11.** Modelo estrutural de avaliação do professor pelo aluno (modelo teórico).

A análise do modelo, utilizando-se o programa AMOS 4.0, método de estimação ML e procedimento de reamostragem, com 500 amostras, forneceu os resultados indicados na Figura 12, em que as cargas fatoriais aparecem numa escala padronizada.



**Figura 12.** Modelo estrutural de avaliação do professor pelo aluno.

Os resultados da análise deste modelo estão na Tabela 20. Os resultados apresentados revelam que doze cargas, representando 75% do total, estão acima de 0,85 e as outras quatro cargas, representando 25% do total estão entre 0,81 e 0,85. Assim, do ponto de vista das cargas fatoriais, o modelo pode ser considerado como bom.

**Tabela 20.** Análise confirmatória do modelo estrutural de avaliação do professor pelo aluno (N=14.951).

variável	carga fatorial	variável	carga fatorial
q03	0,93	q01	0,87
q04	0,91	q13	0,87
q10	0,91	q15	0,86
q02	0,90	q16	0,86
q12	0,90	q07	0,85
q09	0,89	q14	0,85
q06	0,88	q08	0,84
q11	0,88	q05	0,81
índices de adequação			
	CFI	0,958	intervalo de confiança (0,90)
	GFI	0,901	limite inferior      limite superior
	RMSEA	0,088	0,087                  0,089

Quanto aos índices de adequação, o valor do índice RMSEA, no limiar do limite de aceitação, indica uma adequação razoável do modelo. O índice CFI indica uma adequação muito boa e o índice GFI indica uma boa adequação. Podemos, portanto, considerar os índices, no seu conjunto, como satisfatórios.

Podemos, ressaltar, ainda, que o modelo estrutural completo postulou um relacionamento entre o fator relacionamento professor-aluno e o fator prática de ensino. Este relacionamento está ilustrado na Figura 12 por meio de uma correlação de 0,90. Esta correlação, que pode ser considerada alta, pode ser interpretada como responsável pelas modestas cargas (0,48 e 0,54), referentes a cada um dos fatores, isoladamente. Além disso, o valor mais alto para a carga correspondente ao fator prática de ensino é coerente com os resultados relatados na Tabela 13, em que ocorre a predominância, ainda que não exacerbada, do fator prática de ensino sobre o fator relacionamento professor-aluno.

Uma vez analisado o modelo estrutural e obtidos os índices de adequação, uma prática corrente nos estudos que empregam modelagem por equações estruturais consiste na reespecificação e reestimação do modelo, baseado no exame dos índices de modificação do modelo, até a obtenção de um modelo com índices de adequação melhores que os obtidos para o modelo original (Arbuckle & Wothke, 1999; Byrne, 2001; Pasquali, 2006). Nesta tese não empregamos esta abordagem com base em duas razões. Em primeiro lugar, entendemos que este processo de reespecificação e reestimação, com base nos índices de modificação, repetido até a obtenção de um modelo com aceitáveis índices de adequação, é uma análise *post hoc*, que deixa de ser confirmatória e passa a ser de natureza exploratória (Byrne, citado por Cheung, 2000). Em segundo lugar, entendemos que a análise

confirmatória não deve ser guiada por um índice de caráter puramente estatístico ou matemático, mas sim por um sólido conhecimento teórico e substantivo por parte do pesquisador (Byrne, Shavelson & Muthén, citado por Cheung, 2000). Assim, como não há razão substantiva para a reespecificação do modelo, consideramos o modelo estrutural de avaliação do professor pelo aluno, indicado na Figura 11, como adequado.

## Capítulo 6 – Considerações Finais

O sistema educacional brasileiro pode ser visualizado em níveis hierárquicos. No plano superior encontra-se o Ministério da Educação (MEC), responsável pela formulação e implantação de políticas nacionais de educação. A seguir estão as Secretarias Estaduais de educação, responsáveis pela gestão da educação no estado. No terceiro nível encontram-se as Secretarias Municipais de educação, responsáveis pela elaboração e implementação de planos e programas educacionais no âmbito do município. Finalmente, no nível inferior, encontram-se as escolas, tanto públicas quanto privadas, responsáveis pela sua administração, gestão de recursos, relacionamento com a comunidade, implantação das políticas, projetos e programas elaborados pelos três níveis anteriores e, acima de tudo, pela qualidade do ensino oferecido (Fukuda, 2003).

Com base nessa hierarquia, não é difícil concluir que a escola é o ponto de convergência das pressões dos diversos níveis: nacional, regional, municipal e local. Estas pressões são bastante exarcebadas no caso específico do ensino superior, dado o quadro de acirrada concorrência e alta competitividade entre as Instituições de Educação Superior (IES).

Este cenário exige das IES a necessidade de profissionalização de sua gestão, aliando a desburocratização de processos à oferta de serviços ágeis e eficientes. Todo este esforço, no entanto, será inútil, se as IES não conseguirem manter os alunos atuais e conquistar novos, cada vez mais exigentes e com maiores expectativas. Este objetivo, entretanto, só será alcançado, se as IES responderem, com competência e qualidade, a todas essas pressões, tornando-se locais de sucesso, atraentes, desafiadores e convidativos para os alunos (Hattie, 1989, 2003). Essa resposta passa, necessariamente, pela excelência dos processos de ensino e aprendizagem, no âmago dos quais encontramos, além do aluno, um ator fundamental: o professor.

No Brasil observa-se, sem dificuldade, uma deterioração nas condições de vida e de trabalho dos professores, na qualidade e no resultado de sua ação e em sua imagem e auto-estima profissional (Fukuda, 2003). O professor é representante de uma profissão que se encontra extremamente desvalorizada na sociedade, na maior parte das vezes mal remunerada e com péssimas condições de trabalho. O professor é visto pela sociedade, de

uma maneira geral, como uma pessoa que não teve sucesso profissional e, portanto, não tem outras alternativas, senão ensinar, sendo atormentado, a todo instante, pelos infamantes ditados populares do tipo “quem sabe faz, quem não sabe, ensina”.

Apesar dessa desvalorização e de uma preconceituosa, discriminatória e injusta visão do professor, acreditamos que existe uma grande margem de ação, no ensino superior, que é propriamente educativa, e que se desenvolve em sala de aula. Esta ação envolve os dois principais atores dos processos de ensino e aprendizagem: o professor e o aluno. Esses processos podem ser aprimorados, conforme demonstram várias décadas de pesquisa sobre a efetividade do ensino, por meio da avaliação do professor pelo aluno, uma vez que esta avaliação, bem planejada, construída e administrada, contribui de forma significativa para a melhoria da qualidade do ensino.

É neste contexto, de utilização dos resultados obtidos com a avaliação, para o aperfeiçoamento da prática docente e para a elevação da qualidade do ensino nas IES, que se inseriu a proposta da presente tese. Nesta tese, com base nos dados fornecidos pela revisão de literatura, foi desenvolvido um modelo teórico de avaliação, pelo aluno, da prática docente no ensino superior. Além do modelo foi desenvolvido um instrumento, cuja construção foi metodologicamente fundamentada, e cujo objetivo foi o fornecimento de dados para a validação empírica do modelo teórico proposto.

Após três estudos pilotos o instrumento foi aplicado em uma IES do Distrito Federal, envolvendo alunos de 654 turmas e 227 professores. A análise exploratória do banco de dados, levando-se em conta os casos de múltipla marcação de respostas, valores ausentes e valores extremos, resultou numa base de trabalho com 14.951 observações.

O instrumento possui validade de conteúdo, assegurada pela preocupação com a sua construção, com a identificação dos fatores referentes à prática docente e pela sua operacionalização por meio de uma série de comportamentos representativos. A validade de construto, indicativa da estrutura dimensional básica, foi assegurada pela análise fatorial do instrumento, que apresentou uma multifatorialidade compatível com os resultados encontrados na literatura relevante da área.

As análises revelaram a existência de dois fatores: prática de ensino (nove itens) e relacionamento professor-aluno (sete itens). A validade fatorial do instrumento foi assegurada pelo alto valor das cargas fatoriais e pela pureza dos itens: todos os itens foram itens simples e com cargas fatoriais acima de 0,50. Estes resultados mostram, ainda, uma

solidez dos itens que compõem o instrumento, uma vez que todos os fatores possuem mais de cinco itens com cargas fatoriais acima de 0,50 (Costello & Osborne, 2005).

As análises preliminares e confirmatória, utilizando-se, respectivamente, as técnicas de análise fatorial e modelagem por equações estruturais, forneceram dois fatores, relacionamento professor-aluno e prática de ensino, presentes em praticamente todos os estudos relevantes da área. Estes resultados asseguram a invariância fatorial do instrumento. Os altos valores obtidos para o alfa de Cronbach, 0,96 para o fator relacionamento professor-aluno e 0,97 para o fator prática de ensino, asseguram a precisão do instrumento. O instrumento pode ser considerado, então, como um instrumento com boas propriedades psicométricas e como um instrumento adequado para validar o modelo proposto.

Como vimos, o modelo desenvolvido e validado empiricamente é de natureza multifatorial, consoante a literatura relevante sobre a temática abordada, e engloba dois fatores de natureza universal. Estas dimensões de caráter geral, representativas de aspectos relevantes de um ensino de boa qualidade, em sala de aula, e aplicáveis a todos os docentes de uma IES, são a prática de ensino e o relacionamento professor-aluno. Um professor que mantém um bom relacionamento professor-aluno é um professor que, além de manifestar respeito pelo aluno como pessoa, favorece a participação dos alunos, não agride seus alunos e trata a todos com equidade e justiça. Um professor que possui uma boa prática de ensino é um professor claro e preciso nas tarefas propostas, capaz de ministrar as aulas com clareza, autoridade e respeito e capaz de discutir com convicção, relevância e clareza sobre pontos de vista contrários aos seus.

Uma primeira contribuição desta tese está na construção de um modelo de avaliação, pelo aluno, da prática docente no ensino superior. O modelo proposto se destaca em relação aos modelos apresentados na literatura por algumas características.

A primeira característica do modelo proposto é a parcimônia em relação à dimensionalidade. Ele capta, na sua essência, as dimensões universais constituídas por variáveis de caráter geral, representativas de aspectos relevantes de um ensino de boa qualidade, em sala de aula, e aplicáveis a todos os docentes de uma IES. Estas dimensões universais são a prática de ensino e o relacionamento professor-aluno. Uma segunda característica do modelo que deve ser ressaltada como um diferencial é a sua estrutura, conceitualmente construída com base num suporte substantivo teórico adequado. Uma terceira característica peculiar do modelo é que ele é factível de ser implementado de uma

forma rápida e eficiente. Uma quarta característica marcante do modelo é a sua flexibilidade para permitir ampliações graduais e adaptativas. Finalmente o modelo possibilita, no contexto brasileiro, uma melhor compreensão do construto prática docente no ensino superior.

A segunda contribuição desta tese reside na fundamentação metodológica, baseada nas melhores práticas, segundo a literatura, para a construção de um instrumento destinado à validação empírica do modelo desenvolvido.

A terceira contribuição desta tese é o oferecimento de um instrumento, com boas propriedades psicométricas, utilizado para a validação do modelo proposto. Esta validação, com base em dados empíricos, significa que o modelo não foi falseado, no sentido da lógica epistemológica de falsificação de produção de conhecimento (Popper, 1972, 2003).

A quarta contribuição desta tese consiste na metodologia empregada na análise dos dados e na discussão dos resultados obtidos. Estes procedimentos foram realizados com auxílio de análises estatísticas, calibradas, a todo instante, pela relevância dos aspectos teóricos que forneceram o suporte substantivo ao trabalho realizado.

A necessidade desta calibração, no nosso entendimento, reside no fato de que Ciência não é a afirmação de um conjunto de crenças, mas sim um processo de investigação que visa a construção de um corpo testável de conhecimento, constantemente aberto à rejeição ou confirmação (Sherman, citado por Kida, 2007). Desta forma, o conhecimento científico é uma mescla de dúvida e certeza, levando o pesquisador, muitas vezes, a ser arrogantemente humilde: arrogante no método e humilde na fé que tem no seu conhecimento (Bachrach, 1969). Assim, os dados empíricos devem produzir resultados que validem o modelo proposto ou levem o pesquisador a uma reflexão sobre a necessidade de revisão teórica na estrutura conceitual que dá suporte ao modelo. Desta forma o pesquisador deve ter a disposição para aceitar fatos mesmo quando eles se opõem aos desejos, pois, afinal, o erro nada tem de estranho: é o primeiro estado de todo conhecimento (Alain, citado por Lentin, 1997) e, muitas vezes, é melhor ficar sem resposta do que aceitar uma resposta inadequada (Skinner, citado por Bachrach, 1969). Afinal de contas, ausência de evidência não é evidência de ausência, e pesquisas interessantes são resultantes não só de bons planejamentos e execuções, mas também de boas decisões, baseadas em competência técnica, experiência, argumentos sólidos e boas doses de intuição, criatividade e *serendipity*.

A quinta contribuição desta tese apóia-se no nosso entendimento de que apenas desenvolver a crítica não basta para ninguém, uma vez que assim procedendo apenas a tornamos estéril e vazia de referenciais concretos. Precisamos, portanto, aliar aquilo que é necessário saber para saber fazer com o que é necessário saber para analisar, refletir e criticar para transformar. Assim, devemos, com base nos resultados obtidos, fornecer os subsídios para que os professores tenham oportunidade de refletir sobre a melhoria da qualidade de sua prática docente.

Esta oportunidade de reflexão deve ser oferecida, no nosso entendimento, na forma de um *feedback* que, efetivamente, gere ressonâncias sobre o saber, o agir e o ser dos professores. Assim, o *feedback* deve ser fornecido aos professores tão logo seja possível. Os resultados advindos dessa reflexão poderão ser melhor utilizados se mediados por um consultor educacional que seja autêntico, respeitoso, empático, não julgador. O consultor deve, no nosso entendimento, manter uma dinâmica que emerge naturalmente da interação entre ele e o professor, propiciando a este um ambiente psicologicamente seguro para a realização de sua reflexão. O consultor educacional deve estar consciente, entre outras coisas, de que o *feedback* é mais efetivo quando leva em conta o locus de controle e a auto-estima da pessoa a quem o *feedback* se destina (Brinko, 1993). O *feedback* é também mais efetivo quando a informação negativa, que deve ser limitada e selecionada, é oferecida, de forma ensanduichada, entre informações positivas e quando a informação positiva é atribuída a causas internas e a informação negativa é autoreferenciada (Brinko, 1993).

Acreditamos, ainda, que a avaliação dos alunos pode e deve ser analisada juntamente com a auto-avaliação do professor. Eventuais discrepâncias entre essas duas avaliações tendem a criar uma dissonância cognitiva que favorece a reflexão visando a melhoria da qualidade da prática do trabalho em sala de aula e a dissonância cognitiva, em nível moderado, cria um clima psicológico que prepara as pessoas para as mudanças. As eventuais divergências, ao serem bem explicitadas, podem ser compreendidas e serem geradas convergências.

Os professores devem ser sensibilizados e estar preparados para ter a percepção de que os alunos são bem intencionados, confiáveis e capacitados ao realizar a avaliação do trabalho do professor em sala de aula. O professor deve estar consciente de que os dados obtidos não são resultantes de um episódio esporádico, mas sim elementos componentes de um processo que visa a melhoria da qualidade de sua prática docente. Sua reflexão deve

ser focada nos dados concretos fornecidos pela avaliação, devendo estes serem encarados pelo professor como evidências de seu comportamento em sala de aula e não como dados de avaliação de caráter pessoal. Finalmente, acreditamos que esta reflexão só produzirá bons frutos se os professores estiverem conscientes da valorização da relação professor-aluno como instância primordial de uma relação educativa.

Não entendemos esta tese como um estudo exaustivo e nem terminativo. Muitas outras pesquisas podem e devem ser realizadas visando complementar e enriquecer nossa proposta ou mesmo refutá-la à luz de melhores argumentos. A título de sugestão para futuros trabalhos, na temática abordada nesta tese, podemos incluir:

- Ampliação do modelo de avaliação

O modelo de avaliação desenvolvido e validado empiricamente nesta tese, ao privilegiar a parcimônia de dimensões, viabilidade e efetividade de implementação prática, representa apenas um modelo. Ele não é, portanto, o modelo. Ele sugere um arcabouço básico que deve tomar corpo diferenciado em diferentes ambientes e é flexível o suficiente para acomodar mudanças que se mostrem necessárias com o decorrer do tempo. Assim, poderiam ser realizadas pesquisas que ampliem ou sejam alternativas ao modelo proposto.

- Estudos pontuais

Uma sugestão seria incorporar elementos pontuais às análises estatísticas, podendo serem consideradas entre outros: gênero do professor, gênero do aluno, tamanho das classes, titulação do professor, tempo de magistério do professor, etc. Com base nestas variáveis poderiam ser replicados, eventualmente, alguns dos resultados obtidos na literatura e citados no Capítulo 2.

- Ampliação das fontes de informação

Consideramos ser de fundamental importância para as IES a implantação de um programa de melhoria da qualidade do trabalho docente, em que a avaliação do professor pelo aluno seja um elemento importante, porém, não o único componente do programa.

Professores, aqui estamos. Muitos de nós não pediram esse papel e podem não ser talhados para ele. Apesar disso, não podemos renegá-lo. Devemos fazer algo mais que um bom trabalho em sala de aula. Não basta apenas manifestar respeito pelo aluno como pessoa, favorecer a sua participação, não agredi-lo e tratar a todos com equidade e justiça. Não basta apenas sermos claros e preciso nas tarefas propostas, ministrar as aulas com

clareza, autoridade e respeito e sermos capazes de discutir com convicção, relevância e clareza sobre pontos de vista contrários aos nossos.

Apesar de toda as adversidades relativas a condições de trabalho, carreira, salário, apoio técnico, etc, precisamos ser líderes e não vítimas de uma revolução educacional. Líderes que entendam que quem pode mudar uma situação são as pessoas. Líderes que entendam que qualidade passa necessariamente e sobretudo pelas pessoas como seres interacionais. Líderes que compreendam que exercem uma transformação impactante em seus alunos. Líderes que entendam que ser professor é ser um construtor, junto com seus alunos, dia a dia, do conhecimento. Líderes que compreendam que a escola deve inspirar vontade de aprender. Líderes que estejam dispostos a aprender e a encarar questões sob novos pontos de vista. Líderes que entendam que aprender nos capacita a estabelecer prioridades e a encontrar soluções criativas para as mais diversas questões.

Enfim, devemos compreender que a educação, como prática fundamental da existência histórico-cultural dos homens, precisa continuar a ser pensada. Desta forma, compreendendo como os homens construíram sua história no passado, com diversidade de perspectivas, talvez seja possível construir, mediante a práxis atual, a história do futuro. Para que esta história possa ser bem construída, precisamos acreditar na disposição de aprender como uma qualidade humana fundamental. Afinal de contas, talvez melhor que *Cogito, ergo sum* (penso, logo existo) seja *Disco ergo sum* (aprendo, logo existo). Para tornar realidade esta afirmação precisamos acreditar que ser professor é também ser aluno, pois precisamos aprender não só a ensinar, mas sobretudo aprender a aprender.

## Referências

- Abbot, R.D. & Perkins, D. (1978). Development and construct validation of a set of student rating of instruction items. *Educational and Psychological Measurement*, 78, 1069-1075.
- Abrami, P.C., d'Apollonia, S. & Cohen, P.A. (1990). Validity of student ratings of instructions: What we know and what we do not know. *Journal of Educational Psychology*, 82(2), 219-231.
- Afif, A., Clark, V.A. & May, S. (2004). *Computer-aided multivariate analysis (2ª ed.)*. Boca Raton: CRC.
- Aiken, L.R. (1983). Number of response categories and statistics on a teacher rating scale. *Educational and Psychological Measurement*, 43, 397-401.
- Aleamoni, L.M. (1978). Development and factorial validation of the Arizona Course/Instructor Evaluation Questionnaire. *Educational and Psychological Measurement*, 38, 1063-1067.
- Aleamoni, L.M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, 13(2), 153-166.
- Aleamoni, L.M. & Graham, M.H. (1974). The relationship between CEO ratings and instructor's rank, class size, and course level. *Journal of Educational Measurement*, 11, 189-202.
- Alves, L.J.M. & Peres-dos-Santos, L.F.B. (1977). *Matemática 2º grau. Volume 1*. São Paulo: Editora do Brasil.
- Alves, L.J.M. & Peres-dos-Santos, L.F.B. (1980a). *Matemática 2º grau. Volume 2*. São Paulo: Editora do Brasil.
- Alves, L.J.M. & Peres-dos-Santos, L.F.B. (1980b). *Matemática 2º grau. Volume 2. Livro do mestre*. São Paulo: Editora do Brasil.
- Alves-Mazzotti, A.J. & Gewandsznajder, F. (2002). *O método nas ciências naturais e sociais: pesquisa quantitativa e qualitativa*. São Paulo: Pioneira Thomson Learning.
- Amemiya, Y. & Anderson, T.W. (1990). Asymptotic chi-squares tests for a large class of factor analysis models. *Annals of Statistics*, 18, 1453-1463.
- Anscombe, F.J. (1960). Rejection of outliers. *Technometrics*, 2, 123-147.
- Andersen, K. & Miller, E.D. (1997). Gender and student evaluation of teaching. *Political Science and Politics*, 30(2), 216-219.

- Anderson, L.A. & Guichard, D.R. (1988). Statistical effects of class size and teaching ability in determining teacher ratings. *The American Statistician*, 42(1), 55-59.
- Anderson, T.W. (1989). Linear latent variable models and covariance structures. *Journal of Econometrics*, 41, 91-119.
- Anderson, T.W. (2003). *An introduction to multivariate analysis*. New York: Wiley.
- Andrade, J.M. (2005). *Construção de um modelo explicativo de desempenho escolar: Um estudo psicométrico e multinível com dados do SAEB*. Dissertação de Mestrado, Universidade de Brasília, Brasília.
- Arbuckle, J.L. & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago: SPSS.
- Arrindel, W.A. & van der Ende, J. (1985). An empirical-test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, 9(2), 165-178.
- Atamian, R. & Ganguli, G. (1993). Teacher popularity and teaching effectiveness: Viewpoint of accounting students. *Journal of Education for Business*, 68(3), 163-169.
- Babbie, E. (2003). *Métodos de pesquisas de survey*. Belo Horizonte: UFMG.
- Bachrach, A. J. (1969). *Introdução à pesquisa psicológica*. São Paulo: Herder.
- Bain, K. (2004). *What the best college teachers do*. Cambridge: Harvard University Press.
- Baird, J.S. (1987). Perceived learning in relation to student evaluation to university instruction. *Journal of Educational Psychology*, 79(1), 90-91.
- Barnett, V. & Lewis, T. (1994). *Outliers in statistical data (3<sup>a</sup> ed.)*. New York: Wiley.
- Bartholomew, D.J., Steele, F., Moustaki, I. & Galbraith, J.I. (2002). *The analysis and interpretation of multivariate data for social scientists*. Boca Raton: CRC.
- Bartlett, M.S. (1950). Test of significance in fator analysis. *British Journal of Psychology*, 3, 77-85.
- Basow, S.A. (1994). Student ratings of professors are not gender blind. Retirado em 30/11/2005 de <http://www.awm-math.org/newsletter/199409/basow.html>.
- Basow, S.A. & Distenfeld, M.S. (1985). Teacher expressiveness: More important for males than females? *Journal of Educational Psychology*, 77, 45-52.
- Basow, S.A. & Silberg, N.T. (1987). Student evaluation of college professors: Are female and male professors rated differently? *Journal of Educational Psychology*, 79(3), 308-314.

- Bem, A.M. (2004). *Confiabilidade e validade estatísticas da avaliação docente pelo discente: proposta metodológica e estudo de caso*. Tese de Doutorado, Universidade Federal de Santa Catarina, Florianópolis.
- Bendig, A.W. (1954). Reliability and the number of rating scales categories. *Journal of Applied Psychology*, 38, 38-40.
- Bennet, S.K. (1982). Student perception and expectations of male and female instructors: Evidence relating to the questions of gender bias in teaching evaluations. *Journal of Educational Psychology*, 7, 170-179.
- Benson, P.H. (1971). How many scales and how many categories shall we use in consumer research? A comment. *Journal of Marketing*, 35, 59-61.
- Bentler, P.M. & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, 25(1), 67-74.
- Berk, R.A. (1979). The construction of rating instruments for faculty evaluation: A review of methodological issues. *The Journal of Higher Education*, 50(5), 650-669.
- Bernstein, I.H., Garbin, C. & Teng, G. (1988). *Applied multivariate analysis*. New York: Springer-Verlag.
- Boclin, R. (2004). Avaliação de docentes do ensino superior: um estudo de caso. *Ensaio: Avaliação de Políticas Públicas em Educação*, 12(45), 959-980.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling*, 7(3), 461-483.
- Brinko, K.T. (1993). The practice of giving feedback to improve teaching: What is effective? *The Journal of Higher Education*, 64(5), 574-593.
- Browne, M.W. & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, 41, 347-357.
- Bryman, A. & Cramer, D. (2003). *Análise de dados em ciências sociais (3ª ed.)*. Oeiras : Celta.
- Burdsal, C.A. & Bardo, J.W. (1986). Measuring student's perceptions of teaching: Dimensions of evaluation. *Educational and Psychological Measurement*, 46, 63-79.
- Byrne, B.M. (2001). *Structural equation modeling with AMOS: basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.
- Carmine, E. & Zeller, R.A. (1979). *Reliability and validity assessment*. Beverly Hills: Sage.

- Carvalho, H. (2004). *Análise multivariada de dados qualitativos*. Lisboa: Sílabo.
- Cashin, W.E. (1995). *Student ratings of teacher: The research revisited (relatório n.32)*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Centra, J.A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44(5), 495-518.
- Centra, J.A. & Gaubartz, N.B. (2000). Is there gender bias in student evaluation of teaching? *The Journal of Higher Education*, 71(1), 17-33.
- Champney, H. & Marshall, H. (1939). Optimal refinement of the rating scale. *Journal of Applied Psychology*, 23, 323-331.
- Cheung, D. (2000). Evidence of a single second-order factor in student ratings of teaching effectiveness. *Structural Equation Modeling*, 7(3), 442-460.
- Clark, A.C. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309-319.
- Cohen, A. (1983). On the effect of class size on the evaluation of lecturers' performance. *The American Statistician*, 37(4), 331-333.
- Comrey, A.L. & Lee, H.B. (1992). *A first course in factor analysis* (2<sup>a</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Costello, A.B. & Osborne, J.W. (2005). *Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis*. Retirado em 01/02/2007 de <http://PAREonline.net>.
- Costin, F., Greenough, W.T. & Menges, R.J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, 41, 511-535.
- Cox, E.P. III. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, XVII, 407-422.
- Cowan, J. (2002). *Como ser um professor universitário inovador: reflexão na ação*. Porto Alegre: Artmed.
- Cozby, P.C. (2003). *Métodos de pesquisa em ciências do comportamento*. São Paulo: Atlas.
- Crawford, C.B. & Koopman, P. (1973). A note on Horn's test for the number of factors in factor analysis. *Multivariate Behavioral Research*, 8, 117-125.

- Crittenden, K.S., Norr, J.L. & LeBailly, R.K. (1975). Size of university classes and student evaluation of teaching. *The Journal of Higher Education*, 46(4), 461-470.
- Crocker, L. & Algina, J. (1986). *Introduction to classical & modern test theory*. New York: Hartcourt Brace Jovanovich.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 31, 93-96.
- Cunha, M.I. (1989). *O bom professor e sua prática (14ª ed.)*. Campinas: Papyrus.
- Davies, L. & Gather, U. (1993). The identification of multiple outliers (with discussion). *Journal of the American Statistical Association*, 88, 782-792.
- Dillman, D.A. (2000). *Mail and internet surveys: The tailored design method (2ª ed.)*. New York: Wiley .
- Dixon, W.J. (1950). Analysis of extreme values. *Annals of Mathematical Statistics*, 21, 488-506.
- Downie, N.E.W. (1952). Student evaluation of faculty. *Journal of Higher Education*, 23, 495-496, 503.
- Doyle, K.O. & Whitely, S.E. (1974). Student ratings as criteria for effective teaching. *American Educational Research Journal*, 11, 259-274.
- Drucker, A.J. & Remmers, H.H. (1951). Do alumni and students differ in their attitudes toward instructors? *Journal of Educational Psychology*, 42, 129-143.
- Dukes, R.L. & Victoria, G. (1989). The effects of gender, status, and effective teaching on the evaluation of college instructors. *Teaching Sociology*, 17(4), 447-457.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.
- Efron, B. & Tibshirani, R.J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.
- Egan, K. (2002). *A mente educada: os males da educação e a ineficiência educacional das escolas*. Rio de Janeiro: Bertrand Brasil.
- Ellet, C.D., Loup, K.S. & Culross, R.R. (1997). Assessing enhancement of learning, personal learning environment, and student efficacy: Alternative to traditional faculty evaluation in higher education. *Journal of Personnel Evaluation in Education*, 11(2), 167-192.
- Elmore, P.B. & LaPointe, K.A. (1974). Effects of teacher sex and student sex on the evaluation of college instructors. *Journal of Educational Psychology*, 66, 386-389.

- Enzmann, D. (1977). RanEigen: a program to determine the parallel analysis criterion for the number of principal components. *Applied Psychological Measurement*, 21, 232.
- Erthal, T.C. (2001). *Manual de Psicometria (6ª ed.)*. Rio de Janeiro: Jorge Zahar.
- Everitt, B.S. & Dunn, G. (2001). *Applied multivariate data analysis*. London: Arnold.
- Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299.
- Fava, J.L. & Velicer, W.F. (1996). The effect of underextraction in factor and component analysis. *Educational and Psychological Measurement*, 56, 907-922.
- Feldman, K.A. (1993). College students' view of male and female college teachers: Part II, Evidence from students' evaluation of their classroom teachers. *Research in Higher Education*, 34(2), 151-211.
- Ferber, M.A. & Huber, J.A. (1975). Sex of student and instructor: A study of student bias. *The American Journal of Sociology*, 80(4), 949-963.
- Figueiredo, V.L.M. de (2001). *Uma adaptação brasileira do teste de inteligência WISC-III*. Tese de Doutorado, Universidade de Brasília, Brasília.
- Finlayson, A. (2002). *Perguntas que resolvem*. Rio de Janeiro: Campus.
- Fleith, D.S. & Costa Júnior, A. (2005) Métodos de pesquisa em psicologia do desenvolvimento: o que é relevante considerar? Em M.A. Dessen & A.L. Costa Júnior (Orgs.), *A ciência do desenvolvimento humano: tendências atuais e perspectivas futuras* (pp. 37-49). Porto Alegre: Artmed.
- Floyd, F.J. & Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.
- Fowler, F.J. (1988). Design and evaluation of survey questions. Em L. Bickman & D.J. Rog (Eds.), *Handbook of applied social research methods*. Thousand Oakes, CA: Sage.
- Franklin, S.B., Gibson, D.J., Robertson, P.A., Pohlmann, J.T. & Fralish, J.S. (1995). Parallel analysis: A method for determining significant principal components. *Journal of Vegetation Science*, 6, 99-106.
- Freeman, H.R. (1994). Student evaluation of college instructors: Effects of type of course taught, instructor gender and gender role, and student gender. *Journal of Educational Psychology*, 86(4), 627-630.
- Fukuda, C. C. (2003). *O ensino eficaz na educação básica: um modelo descritivo dos fatores de eficácia*. Tese de Doutorado, Universidade de Brasília, Brasília.

- Gadotti, M. (2003). *História das idéias pedagógicas*. São Paulo: Ática.
- Gage, N.L. (1961). The appraisal of college teacher. *Journal of Higher Education*, 32, 17-22.
- Garland, R. (1991). The mid-point on a rating scale: is it desirable? *Marketing Bulletin*, 2, 66-70.
- Gaudagnoli, E. & Velicer, W.F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103, 265-275.
- Gigliotti, R.J. & Buchtel, F.S. (1990). Attributional bias and course evaluation. *Journal of Educational Psychology*, 82(2), 341-351
- Goodwin, L.D. & Stevens, E.A. (1993). The influence of gender on university faculty members' perception of "good" teaching. *Journal of Higher Education*, 64(2), 166-185.
- Goldberg, G. & Callahan, J. (1991). Objectivity of student evaluation of instructors. *Journal of Education for Business*, 66(6), 377-378.
- Gorsuch, R.L. (1983). *Factor analysis (2<sup>a</sup> ed.)*. Hillsdale: Lawrence Earlbaum.
- Gorsuch, R.L. (1990). Common factor-analysis versus components analysis- some well and little known facts. *Multivariate Behavioral Research*, 25(1), 33-39.
- Grimm, L.G. & Yarnold, P.R. (Orgs.) (2002). *Reading and understanding more multivariate statistics*. Washington: APA.
- Grimm, L.G. & Yarnold, P.R. (Orgs.) (2003). *Reading and understanding multivariate statistics*. Washington: APA.
- Gunther, H. (2003). *Como elaborar um questionário*. (Série: Planejamento da Pesquisa nas Ciências Sociais, Nº 01). Brasília, DF:UnB, Laboratório de Psicologia Ambiental.
- Guthrie, E.R. (1949). The evaluation of teaching. *Educational Record*, 30, 109-115.
- Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, 19, 149-162.
- Haemmerlie, F.M. & Highfill, L.A. (1991). Bias by male engineering undergraduates in their evaluation of teaching. *Psychological Reports*, 68, 151-160.
- Hair, J.F., Babib, B., Money, A.H. & Samouel, P. (2005). *Fundamentos de métodos de pesquisa em administração*. Porto Alegre: Bookman.
- Hamilton, L.C. (1992). *Regression with graphics: A second course in applied statistics*. Monterey: Brooks/Cole.
- Harlow, L.L. (2005). *The essence of multivariate thinking*. London: Lawrence Erlbaum.

- Harman, H.H. (1976). *Modern factor analysis (3<sup>a</sup> ed.)*. Chicago: University of Chicago.
- Harris, M.B. (1975). Sex roles stereotypes and teacher evaluations. *Journal of Educational Psychology*, 67, 751-756.
- Harrison, P.D., Ryan, J.M. & Moore, P.S. (1996). College students' self-insight and common implicit theories in ratings of teaching effectiveness. *Journal of Educational Psychology*, 88(4), 775-782.
- Hartwig, F. & Dearing, B.E. (1979). *Exploratory data analysis*. Newbury Park : Sage.
- Hativa, N. (1996). University instructors' rating profiles: Stability over time, and disciplinary differences. *Research in Higher Education*, 37(3), 341-365.
- Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139-164.
- Hattie, J. (1989). *The ideal school*. Retirado em 5/09/2006 de <http://www.ats.auckland.ac.nz/staff/index.cfm>.
- Hattie, J. (2003). *Teachers make a difference: what is the research evidence?* Retirado em 5/09/2006 de <http://www.ats.auckland.ac.nz/staff/index.cfm>.
- Hawkins, D.M. (1980). *Identification of outliers*. London: Chapman and Hall.
- Hayes, J.R. (1971). Research, teaching and faculty fate. *Science*, 172, 227-230.
- Hayton, J.C., Allen, D.G. & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191-205.
- Heilman, J.D. & Armentrout, W.D. (1936). The rating of college teachers on ten traits by their students. *Journal of Educational Psychology*, 27, 197-216.
- Hildebrand, M. (1972). How to recommend promotion for a mediocre teacher without actually lying. *The Journal of Higher Education*, 43(1), 44-62.
- Hogan, T.P. (1973). Similarity of student rating across instructors, courses and time. *Research in Higher Education*, 1, 149-154.
- Hogan, T.P. (2006). *Introdução à prática de testes psicológicos*. Rio de Janeiro: LTC.
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Howard, G.S. & Maxwell, S.E. (1980). Correlation between student satisfaction and grades: Case of mistaken causation? *Journal of Educational Psychology*, 72(6), 810-820.

- Hubbard, R. & Allan, S.J. (1987). An empirical comparison of alternative methods for principal component extraction. *Journal of Business Research*, 15, 173-190.
- Huck, S.W. (2000). *Reading statistics and research (3ª ed.)*. New York: Longman.
- Humphreys, L.G. & Ilgen, D.R. (1969). Note on the criterion for the number of common factors. *Educational and Psychological Measurement*, 29, 571-578.
- Humphreys, L.G. & Montanelli, R.G. (1975). Latent roots of random matrix correlation matrices with squared multiple correlations in the diagonals: a Monte Carlo study. *Psychometrika*, 41, 341-348.
- Isaacson, R.L., McKeachie, W.J., Millholland, J.E., Lin, Y.G., Hotelier, M., Baerwaldt, J.W. & Zinn, K.L. (1964). Dimensions of student evaluation of teaching. *Journal of Educational Psychology*, 55, 344-351.
- Jacoby, J. & Matell, M.S. (1971). Three-point Likert scales are good enough. *Journal of Marketing Research*, VIII, 495-500.
- Jesus, G. R. de (2004). *Fatores que afetam o desempenho em português: um estudo multinível com dados do SAEB 2001*. Dissertação de Mestrado, Universidade de Brasília, Brasília.
- Jesus, G.R. de & Laros, J.A. (2004). Eficácia escolar: regressão multinível com dados de avaliação em larga escala. *Avaliação Psicológica*, 3(2), 21-31.
- Judd, C.M. & McClelland, G.H. (1989). *Data analysis: A model comparison approach*. San Diego: Harcourt Brace Jovanovich.
- Kachigan, S.K. (1991). *Multivariate statistical analysis (2ª ed.)*. New York: Radius.
- Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kaplan, A. (1972). *A conduta na pesquisa: metodologia para as ciências do comportamento*. São Paulo: Herder.
- Kaplan, D. (2000). *Structural equation modeling: Foundation and extensions*. Thousand Oaks: Sage.
- Kerlinger, F.N. (1980). *Metodologia da pesquisa em ciências sociais*. São Paulo: EPU/EDUSP.
- Kida, T. (2007). *Não acredite em tudo que você pensa*. Rio de Janeiro: Elsevier.
- Kierstad, D., D'Agostino, P. & Dill, H. (1988). Sex role stereotyping of college professors: bias in student ratings of instructors. *Journal of Educational Psychology*, 80, 342-344.

- Kline, R.B. (1998). *Principles and practices of structural equation modeling*. New York: Guilford.
- Koon, J. & Murray, H.G. (1995). Using multiple outcomes to validate student ratings of overall teacher effectiveness. *Journal of Higher Education*, 66(1), 61-81.
- Laros, J.A. (2005). O uso de análise fatorial: algumas diretrizes para pesquisadores. Em Luiz Pasquali (Org.), *Análise fatorial para pesquisadores* (pp. 163-184). Brasília: LabPAM.
- Laros, J.A. & Puente-Palacios, K.E. (2004). Validação cruzada de uma escala de clima organizacional. *Estudos de Psicologia (Natal)*, 9(1), 113-119.
- Lattin, J., Carrol, J.D. & Green, P.E. (2003). *Analyzing multivariate data*. Pacific Grove: Brooks/Cole.
- Lentin, J. (1997). *Penso, logo me engano*. São Paulo: Ática.
- Linn, R.L. (1968). A Monte Carlo approach to the number of factors problems. *Psychometrika*, 33, 37-71.
- Linsky, A.S. & Strauss, M.A. (1975). Student evaluations, research productivity and eminence of college faculty. *Journal of Higher Education*, 46, 89-102.
- Lissitz, R.W. & Green, S.B. (1975). Effect of the number of scale points on reliability: An Monte Carlo approach. *Journal of Applied Psychology*, 60(1), 10-13.
- Little, R.J. & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Loehlin, J.C. (1990). Component analysis versus common factor-analysis: A case of disputed authorship. *Multivariate Behavioral Research*, 25(1), 29-31.
- Lord, F.M. (1953). On the statistical treatment of football numbers. *The American Psychologist*, 8, 750-751.
- Lowman, J. (2004). *Dominando as técnicas de ensino*. São Paulo: Atlas.
- Ludbrook, J. & Dudley, H. (1998). Why permutation tests are superior to t- and F-tests in biomedical research. *American Statistician*, 52, 127-132.
- Ludwig, J.M. & Meacham, J.A. (1997). Teaching controversial courses: Student evaluation of instructor and contents. *Educational Research Quarterly*, 21(1), 27-38.
- MacCallum, R.C. & Tucker, L.R. (1991). Representing sources of error in the common-factor model – implications for theory and practice. *Psychological Bulletin*, 109(3), 502-511.
- MacCallum, R.C., Widaman, K.F., Zhang, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99.

- Macedo, S.G. (2001). *Desempenho docente pela avaliação discente: uma proposta metodológica para subsidiar a gestão universitária*. Tese de Doutorado, Universidade Federal de Santa Catarina, Florianópolis.
- Malhotra, N.K. (2004). *Pesquisa de marketing: uma orientação aplicada*. Porto Alegre: Bookman.
- Marlin, J.W. (1987). Students perceptions of end-of-course evaluation. *Journal of Higher Education*, 58(6), 704-716.
- Marsh, H.W. (1977). The validity of students' evaluation: classroom evaluation of instructors independently nominated as best and worst teachers by graduating seniors. *American Educational Research Journal*, 14, 441-447.
- Marsh, H.W. (1981). Students' evaluation of tertiary instruction: Testing the applicability of American surveys in an Australian setting. *Australian Journal of Education*, 25, 177-192.
- Marsh, H.W. (1982a). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluation of university teaching. *British Journal of Educational Psychology*, 52, 77-95.
- Marsh, H.W. (1982b). Validity of students' evaluation of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology*, 74, 264-279.
- Marsh, H.W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/ course/ instructor characteristics. *Journal of Educational Psychology*, 75(1), 150-166.
- Marsh, H.W. (1984). Students' evaluation of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707-754.
- Marsh, H.W. (1986). Applicability paradigm: Students' evaluation of teaching effectiveness in different countries. *Journal of Educational Psychology*, 78(6), 465-473.
- Marsh, H.W. (1987). Students' evaluation of university teaching: Research findings, methodological issues, and directions for future research. *Instructional Evaluation*, 10, 5-9.
- Marsh, H.W. (1991a). A multidimensional perspective on students' evaluation of teaching effectiveness: Reply to Abrami and d'Apolonia (1991). *Journal of Educational Psychology*, 83, 416-421.
- Marsh, H.W. (1991b). Multidimensional students' evaluations of teaching effectiveness: A test of alternative higher order structures. *Journal of Educational Psychology*, 83(2), 285-296.

- Marsh, H.W. (1993). Multidimensional students' evaluations of teaching effectiveness. *Journal of Higher Education*, 64(1), 1-18.
- Marsh, H.W. & Bailey, M. (1993). Multidimensional students' evaluation of teaching effectiveness: A profile analysis. *The Journal of Higher Education*, 64(1), 1-18.
- Marsh, H.W. & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement*, 20, 231-248.
- Marsh, H.W. & Hocevar, D. (1984). The factorial invariance of student evaluation of college teaching. *American Educational Research Journal*, 21, 341-366.
- Marsh, H.W. & Hocevar, D. (1991). The multidimensionality of students' evaluation of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. *Teaching & Teacher Education*, 7, 9-18.
- Marsh, H.W. & Overall, J.U. (1979). Long-term stability of students's evaluations: A note on Feldman's consistency and variability among college students in rating their teachers and courses. *Research in Higher Education*, 10, 139-147.
- Marsh, H.W., Overall, J.U. & Kesler, S.P. (1979a). Class size, students' evaluations, and instructional effectiveness. *American Educational Research Journal*, 16(1), 57-69.
- Marsh, H.W., Overall, J.U. & Kesler, S.P. (1979b). Validity of student evaluation of instructional effectiveness: A comparison of faculty self-evaluations and evaluations by their students. *Journal of Educational Psychology*, 71(2), 149-160.
- Marsh, H.W. & Roche, L.A. (1997). Making students' evaluation of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), 1187-1197.
- Maslow, A.H. & Zimmermann, W. (1956). College teaching ability, scholarly activity and personality. *Journal of Educational Psychology*, 47, 185-189.
- Masetto, M.T. (2003). *Competência pedagógica do professor universitário*. São Paulo: Summus.
- Matell, M.S. & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert-scale items? Study I: Reliability and validity. *Educational and Psychological Measurement*, 31, 657-674.
- Matell, M.S. & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*, 56(6), 506-509.
- Mateo, M.A. & Fernandez, J. (1996). Incidence of class size on the evaluation of university teaching quality. *Educational and Psychological Measurement*, 56(5), 771-778.

- McArdle, J.J. (1990). Principles versus principals of structural factor-analyses. *Multivariate Behavioral Research*, 25(1), 81-87.
- McArdle, J.J. & McDonald, R.P. (1984). Some algebraic properties of the reticular model for moment structures. *British Journal of Mathematical and Statistical Psychology*, 37, 234-251.
- McDonald, R.P. & Ho, M.R. (2002). Principles and practice in reporting structural equations analyses. *Psychological Methods*, 7(1), 64-82.
- McGrath, E.J. (1962). Characteristics of outstanding college teachers. *Journal of Higher Education*, 33, 148.
- McGuigan, F.J. (1976). *Psicologia experimental: uma abordagem metodológica*. São Paulo: EPU.
- McKeachie, W.J. (1979). Student ratings of faculty: A reprise. *Academe*, 65, 384-397.
- Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, 100, 398-407.
- Miles, J. & Shevlin, M. (2001). *Applying regression & correlation*. London: Sage.
- Miller, R. I. (1972). *Evaluating faculty performance*. San Francisco: Jossey-Bass.
- Millman, J. (Ed.) (1981). *Handbook of teacher evaluation*. Beverly Hills: Sage.
- Mingoti, S. A. (2005). *Análise de dados através de métodos de estatística multivariada*. Belo Horizonte: UFMG.
- Ministério da Educação e Cultura. (2001). *Educação no Brasil: 1995-2001*. Brasília: Autor.
- Morais, R. (1986). *O que é ensinar*. São Paulo: EPU.
- Morais, R. (Org.). (1988). *Sala de aula: que espaço é esse? (17ª ed.)*. Campinas: Papirus.
- Morales, P. (2001). *A relação professor-aluno: o que é, como se faz (3ª ed.)*. São Paulo: Loyola.
- Moreira, D.A. (1986). *Avaliação do professor universitário pelo aluno: possibilidades e limitações*. Tese de Doutorado, Universidade de São Paulo, São Paulo.
- Moreira, D.A. (1988). Fatores influentes na avaliação do professor: uma revisão. *Educação e Seleção*, 17, 73-87.
- Moritsch, B.G. & Suter, W.N. (1988). Correlates of halo error in teacher evaluation. *Educational Research Quarterly*, 12(3), 29-34.

- Newton, R.R. & Rudestam, K.E. (1999). *Your statistical consultant: Answers to your data analysis questions*. Thousand Oaks, CA: Sage.
- Nikolic, V. & Cabaj, H. (2001). *Estou ensinando bem? Estratégias de auto-avaliação para professores*. São Paulo: Loyola.
- Nimmer, J.G. & Stone, E.F. (1991). Effects of grading practices and time of rating on student ratings of faculty performance and student learning. *Research in Higher Education*, 32(2), 195-215.
- Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric theory* (3<sup>a</sup> ed.). New York: McGraw-Hill.
- O'Connor, B.P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers*, 32, 396-402.
- O'Connor, B.P. (2006). *Cautions regarding item-level factor analyses*. Retirado em 14/4/2006 de <http://flash.lakeheadu.ca/~boconno2/itemanalysis.html>.
- Oliveira, T.M.V. (2001). Escalas de mensuração de atitudes: Thurstone, Osgood, Stapel, Likert, Guttman, Alpert. *Administração On Line*, 2(2), 1-25. Retirado em 6/5/2004 de [http://www.fecap.br/adm\\_online/art22/tania.htm](http://www.fecap.br/adm_online/art22/tania.htm).
- Osborne, J.W. (2002). *Notes on the use of data transformations*. Retirado em 30/03/2005 de <http://PAREonline.net>.
- Osborne, J.W. & Overbay, A. (2004). *The power of outliers*. Retirado em 01/02/2007 de <http://PAREonline.net>.
- Pasquali, L. (1984). Questionário de avaliação de docência (QAD). *Educação e Seleção*, 9, 71-98.
- Pasquali, L. (Org.). (1999). *Instrumentos psicológicos: manual prático de elaboração*. Brasília: LabPAM/IBAPP.
- Pasquali, L. (Org.). (2001). *Técnicas de exame psicológico - TEP- Manual. Volume I*. São Paulo: Casa do Psicólogo.
- Pasquali, L. (2003). *Psicometria: teoria dos testes na Psicologia e na Educação*. Petrópolis: Vozes.
- Pasquali, L. (2004). *Delineamento de pesquisa em ciência*. Brasília: LabPAM.
- Pasquali, L. (Org.). (2005). *Análise fatorial para pesquisadores*. Brasília: LabPAM.
- Pasquali, L. (2006). *Métodos inferenciais avançados*. Brasília: LabPAM.

- Pereira, M.E.M. & Morais, J.C. (2000). *Avaliação na graduação: estudo de caso*. Retirado em 05/12/2005 de <http://www.prg.ufpb.br>.
- Peres-dos-Santos, L.F.B. (1969). *Matemática III*. Brasília: Brasil Central.
- Peres-dos-Santos, L.F.B. (1975). *Matemática básica*. Brasília: Carajás.
- Peres-dos-Santos, L.F.B. (1976a). *Capacidade vital entre estudantes não fumantes e de vida sedentária*. Tema, Segundo Congresso Brasileiro de Pneumologia e Tisiologia, Salvador, BA.
- Peres-dos-Santos, L.F.B. (1976b). *Capacidade vital entre estudantes não fumantes e de vida sedentária*. Tema, Terceira Jornada Internacional de Pneumologia, Salvador, BA.
- Peres-dos-Santos, L.F.B. (1978). *Convergência de algoritmos de otimização*. Dissertação de Mestrado, Universidade de Brasília, Brasília.
- Peres-dos-Santos, L.F.B. (1980). Cuidado com a estatística e com os estatísticos. *Boletim da Escola Superior de Administração Postal*, 3(6), 11-13.
- Peres-dos-Santos, L.F.B. (1981). A análise estatística a serviço do administrador. *Leia*, 1(4), 21-24.
- Peres-dos-Santos, L.F.B. (1987). *Segurança em correio eletrônico*. Dissertação de Mestrado, Instituto Militar de Engenharia, Rio de Janeiro.
- Peres-dos-Santos, L.F.B. & Laros, J.A. (2007). Avaliação da prática pedagógica do professor de ensino superior. *Estudos em Avaliação Educacional*, 18(36), 75-95.
- Pimenta, S.G. & Anastasiou, L.G.C. (2002). *Docência no ensino superior. Volume I*. São Paulo: Cortez.
- Pimentel, M.G. (1993). *O professor em construção (5ª ed.)*. Campinas: Papyrus.
- Pinent, C.E.C., Silveira, F.L. & Moraes, R. (1993). Avaliação do professor pelo aluno: questionário avaliativo e testes de validação. *Estudos em Avaliação Educacional*, 8, 65-79.
- Popper, K. (1972). *A lógica da pesquisa científica*. São Paulo: Cultrix.
- Popper, K. (2003). *Conjecturas e refutações*. Coimbra: Almedina.
- Posavac, E.J., Carey, R.G. (2003). *Program evaluation methods and case studies (6ª ed.)*. Upper Saddle River, NJ: Sage.
- Prave, R.S. & Baril, G.L. (1993). Instructor ratings: controlling for bias from initial student interest. *Journal of Education for Business*, 68(6), 362-366.

- Prosser, M. & Trigwell, K. (1990). Student evaluation of teaching and courses: student study strategies as a criterion of validity. *Higher Education*, 20(2), 135-142.
- Rangel, M. (2002). *Representações e reflexões sobre o "bom professor" (6ª ed)*. Petrópolis: Vozes.
- Rasmussen, J.L. (1988). Evaluating outliers identification tests: Mahalanobis D squared and Comrey D. *Multivariate Behavioral Research*, 23(2), 189-202.
- Reis, E. (2001). *Estatística multivariada aplicada (2ª ed)*. Lisboa: Sílabo.
- Reise, S.P., Waller, N.G. & Comrey, A.L. (2000). Factor analysis and scale revision. *Psychological Assessment*, 12, 287-297.
- Rios, T.A. (1993). *Ética e competência (11ª ed)*. São Paulo: Cortez.
- Rios, T.A. (2002). *Compreender e ensinar: por uma docência de melhor qualidade (3ª ed)*. São Paulo: Cortez.
- Rousseeuw, P. & Leroy, A. (1987). *Robust regression and outlier detection*. New York: Wiley.
- Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Ryans, D.G. (1960). *Characteristics of teachers*. Washington: American Council on Education.
- Sailor, P., Whorten, B. R. & Shin, E.H. (1997). Class level as a possible mediator of the relationship between grades and student ratings of teaching. *Assesment and Evaluation in Higher Education*, 22(3), 261-269.
- Sant'anna, I.M. (2002). *Por que avaliar? Como avaliar? Critérios e instrumentos (5ª ed)*. Petrópolis: Vozes.
- Santos, J.R. (2004). *Avaliação do desempenho dos docentes e dos cursos da Faculdade Estadual de Ciências Econômicas de Apucarana – FECEA – um estudo de caso*. Dissertação de Mestrado, Universidade Federal de Santa Catarina, Florianópolis.
- Schmelkin, L.P. & Spencer, K.J.(1997). Faculty perspectives on course and teacher evaluation. *Research in Higher Education*, 38(5), 75-92.
- Scherr, F.C. & Scherr, S.S. (1990). Bias in student evaluation of teacher effectiveness. *Journal of Education for Business*, 65(8), 356-358.
- Schwager, S.J. & Margolin, B.H. (1982). Detection of multivariate outliers. *The annals of statistics*, 10, 943-954.
- Scriven, M. (1995). *Student ratings offer useful input to teacher evaluations*. Retirado em 30/03/2005 de <http://PAREonline.net>.

- Shapiro, E.G. (1990). Effect of instructor and class characteristics on students' class evaluations. *Research in Higher Education*, 3(2), 135-148.
- Sidanius, J. & Crane, M. (1989). Job evaluation and gender: The case of university faculty. *Journal of Applied Social Psychology*, 19, 174-197.
- Smith, G.T. & McCarthy, D.M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment*, 7, 300-308.
- Snook, S.C. & Gorsuch, R.L. (1989). Component analysis versus common factor-analysis – a Monte Carlo study. *Psychological Bulletin*, 106(1), 148-154.
- Souza, P.R. (2005). *A revolução gerenciada: Educação no Brasil 1995-2002*. São Paulo: Prentice-Hall.
- Steiger, J.H. (1990). Some additional thoughts on components, factors and factor-indeterminacy. *Multivariate Behavioral Research*, 25(1), 41-45.
- Tabachnick, B.G. & Fidell, L.S. (2001). *Using multivariate statistics (4<sup>a</sup> ed.)*. Needham Heights: Allyn & Bacon.
- Tacq, J. (1997). *Multivariate analysis techniques in social science research*. Thousand Oaks: SAGE.
- Tardif, M. (2002). *Saberes docentes & formação profissional*. Petrópolis: Vozes.
- Tatro, C.N. (1995). Gender effects of student evaluations of faculty. *Journal of Research and Development in Education*, 28, 169-173.
- Tollefson, N., Cheng, J.S. & Kleinsasser, A. (1989). The relationship of perceived teacher caring with student learning and teacher evaluation. *Educational and Psychological Measurement*, 49(3), 529-536.
- Townsend, J.T. & Ashby, F.G. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin*, 96(2), 394-401.
- Trigueiro, M.G.S. (2000). *O ensino superior privado no Brasil*. Brasília: Paralelo 15.
- Ullmann, J.B. (2001). Structural equation modeling. Em: B.G.Tabachnick & L.S. Fidell (Orgs.), *Using multivariate statistics (4<sup>a</sup> ed.)* (pp. 653-771). Needham Heights, MA: Allyn & Bacon.
- Unesco. (2004). *O perfil dos professores brasileiros: o que fazem, o que pensam, o que almejam*. São Paulo: Moderna.
- Valério, R.N. (2004). *Avaliação institucional: Uma relação entre avaliação docente e discente - um estudo de caso*. Dissertação de Mestrado, Universidade Federal de Santa Catarina, Florianópolis.

- Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
- Velicer, W.F. & Jackson, D.N. (1990). Component analysis versus common factor-analysis – some further observations. *Multivariate Behavioral Research*, 25(1), 97-114.
- Vianna, H.M. (2003). *Avaliações em debate: SAEB, ENEM, PROVAO*. Brasília: Plano.
- Voeks, V.W. (1962). Publications and teaching effectiveness. *Journal of Higher Education*, 33, 212.
- Wainer, H. (1976). Robust statistics: A survey and some prescriptions. *Journal of Educational Statistics*, 1(4), 285-312.
- Ware, J.E. & Williams, R.G. (1977). Discriminate analysis of student ratings as a means of identifying lecturers who differ in enthusiasm or information giving. *Educational and Psychological Measurement*, 37, 627-639.
- Waters, M., Kemp, E. & Pucci, A. (1988). High and low faculty evaluations: Descriptions by students. *Teaching of Psychology*, 15(4), 203-204.
- Watkins, D. (1990). Student ratings of tertiary courses for “alternative calendars” purposes. *Assessment and Evaluation in Higher Education*, 15(1), 12-21.
- Widaman, K.F. (1990). Bias in pattern loadings represented by common factor-analysis and component analysis. *Multivariate Behavioral Research*, 25(1), 89-95.
- Widaman, K.F. (1993). Common factor-analysis versus principal component analysis – differential bias in representing model parameters. *Multivariate Behavioral Research*, 28(3), 263-311.
- Wilcox, R.R. (1992). An improved method for comparing variances when distributions have nonidentical shapes. *Computational Statistics & Data Analysis*, 13, 163-172.
- Wilcox, R.R. (1993). Comparing one-step M-estimators of location when there are more than two groups. *Psychometrika*, 58, 71-78.
- Wilcox, R.R. (1994). A one-way random-effects model for trimmed means. *Psychometrika*, 59, 289-306.
- Wilcox, R.R. (1996). *Statistics for the social sciences*. San Diego, CA: Academic Press.
- Wilcox, R.R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.
- Wilcox, R.R. (2003). *Applying contemporary statistical techniques*. San Diego, CA: Academic Press.

- Wilson, R. (1998). New research casts doubt on value of comparing adult college student perceptions of effective teaching with those of traditional students. *Chronicles of Higher Education*, 44(19), A12-A14.
- Wholey, J.S., Hatry, H.P. & Newcomer, K.E. (2004). *Handbook of practical program evaluation (2<sup>a</sup> ed)*. San Francisco, CA: Jossey-Bass.
- Wolins, L. (1995). A Monte Carlo study of constrained factor analysis using maximum likelihood and unweighted least squares. *Educational and Psychological Measurement*, 55, 545-557.
- Worthen, B.R., Sanders, J.R., Fitzpatrick, J.L. (2004). *Avaliação de programas: concepções e práticas*. São Paulo: Gente.
- Wotruba, T.R. & Wright, P.L. (1975). How to develop a teacher-rating instrument: A research approach. *The Journal of Higher Education*, 46(6), 653-663.
- Zwick, R. & Velicer, W.F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

**Anexo 1: Características dos membros do comitê revisor**

- 1 professor de Didática e Psicologia  
Graduação em Pedagogia e em Psicologia  
Mestrado em Educação  
13 anos de experiência de magistério no ensino superior  
6 anos de experiência em avaliação de professores de ensino superior
- 1 professor de Estatística  
Graduação em Contabilidade  
Mestrado em Estatística  
6 anos de experiência de magistério no ensino superior  
6 anos de experiência em avaliação de professores de ensino superior
- 1 professor de Matemática e Estatística  
Graduação em Matemática  
Mestrado em Estatística  
7 anos de experiência de magistério no ensino superior  
3 anos de experiência em avaliação de professores de ensino superior
- 1 professor de Administração  
Graduação em Administração  
15 anos de experiência de magistério no ensino superior  
6 anos de experiência em avaliação de professores de ensino superior
- 1 professor de Matemática e Estatística  
Graduação em Matemática e em Estatística  
Mestrado em Estatística e em Computação  
15 anos de experiência de magistério no ensino superior  
9 anos de experiência em avaliação de professores de ensino superior
- 1 dirigente de IES  
Graduação em Matemática  
18 anos de experiência de magistério no ensino superior  
32 anos de experiência em avaliação de professores de ensino superior, gestão pedagógica e administrativa de IES
- 1 professor de Língua Portuguesa  
Graduação em Língua Portuguesa  
Mestrado em Linguística  
26 anos de experiência de magistério no ensino superior

**Anexo 2: Modelo do questionário aplicado**

## PESQUISA DE OPINIÃO SOBRE O CORPO DOCENTE

### 1º Semestre / 2005

**Caro(a) Aluno(a),**

sua participação no processo de melhoria da qualidade de ensino do xxxxx é muito importante. Responda com atenção, ética, seriedade e coerência os itens abaixo, de acordo com a avaliação da atividade do professor. Agradecemos a colaboração e lhe asseguramos que as informações coletadas serão utilizadas com vista ao enriquecimento da proposta de ensino que oferecemos.

#### INSTRUÇÕES PARA O PREENCHIMENTO

◆ Para cada item, assinale apenas uma alternativa de resposta

◆ Marque assim ➤ ●

#### PRÁTICA PEDAGÓGICA

	Ruim					Bom					Ótimo									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
01 - O professor responde às perguntas e aos comentários dos alunos de forma educada e sem aparentar irritação.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
02 - O professor promove a interação entre os alunos e encoraja o respeito mútuo.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
03 - O professor procura criar uma atmosfera agradável na sala de aula.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
04 - O professor mantém relacionamento cordial com os alunos.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
05 - O professor aplica as mesmas regras para todos, sem favorecer alguns alunos em relação a outros.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
06 - O professor resolve, com habilidade, os problemas de indisciplina em sala de aula.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
07 - O professor ministra as aulas com entusiasmo e bom humor.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
08 - O professor propõe tarefas interessantes para favorecer o aprendizado.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
09 - O professor usa uma linguagem clara para apresentar e desenvolver a matéria.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
10 - O professor responde aos questionamentos dos alunos de forma clara e objetiva.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
11 - O professor orienta, com clareza, os trabalhos propostos.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
12 - O professor desenvolve a matéria num ritmo satisfatório.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
13 - O professor encoraja os alunos a fazer questionamentos, expressando livremente suas idéias.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
14 - O professor propõe provas que valorizam a reflexão mais do que a memorização.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
15 - O professor demonstra coerência entre as avaliações e os conteúdos desenvolvidos em sala de aula.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
16 - O professor utiliza o tempo de aula de forma produtiva.	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10

#### NOTA GLOBAL

	Ruim					Bom					Ótimo									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
17 - A nota que você atribui para o Professor é...	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10

#### DESEMPENHO

	Ruim	Bom	Ótimo
18 - Como você avalia o Professor em relação aos demais professores deste semestre...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

#### CUMPRIMENTO DE NORMAS

	Sim	Não
19 - O professor esclarece, no início do semestre, os objetivos, o conteúdo programático, a metodologia de trabalho e os critérios de avaliação na disciplina.	<input type="radio"/>	<input type="radio"/>
20 - O professor devolve aos alunos no prazo estipulado ( 7 dias para as disciplinas de 4 créditos e 10 dias para as disciplinas de 2 créditos) as provas e trabalhos corrigidos.	<input type="radio"/>	<input type="radio"/>
21 - O professor adota outros procedimentos de avaliação que não sejam apenas provas.	<input type="radio"/>	<input type="radio"/>
22 - O professor cumpre, com pontualidade, os horários de início e término das aulas.	<input type="radio"/>	<input type="radio"/>

#### DADOS DO ALUNO

23 - Curso em que está matriculado

Administração Matutino
  Administração Noturno
  Contabilidade
  Direito Matutino
  Direito Noturno
  Economia
  Pedagogia

24 - Sexo

Feminino
  Masculino

**CONTINUA NO VERSO** ➔

**Anexo 3: Modelos dos e-mails enviados aos alunos e professores da IES**

## E-MAIL ENVIADO AOS ALUNOS

Caro aluno:

No período de 16 a 21 de maio, estaremos realizando a pesquisa de opinião dos alunos sobre os professores.

Trata-se de pesquisa **voluntária**, porém esperamos que você dedique um pouco do seu tempo para oferecer contribuições que julgue adequadas, uma vez que a opinião **séria, ética e coerente** dos alunos é fator essencial para a melhoria da qualidade do ensino oferecido pelo xxxx.

A pesquisa, a ser **realizada em sala de aula**, consta de um formulário contendo alguns itens para os quais você deverá assinalar a opção de resposta que corresponde à sua avaliação sobre a atividade do professor. Em cada horário você emitirá a sua opinião sobre o trabalho desenvolvido pelo professor da respectiva disciplina.

Para garantir a confidencialidade de suas respostas, o professor se ausentará da sala de aula e a aplicação da pesquisa será conduzida por um aluno.

As suas opiniões só serão entregues aos professores correspondentes e aos respectivos coordenadores para uma reflexão conjunta, **depois** que as pautas das disciplinas estiverem encerradas, isto é, estiverem lançadas no sistema as notas e as frequências dos alunos. Desta forma, você terá a garantia de que o seu desempenho na disciplina será independente das opiniões emitidas.

EMITA, PORTANTO, SUAS OPINIÕES COM TRANQUILIDADE, SERIEDADE, COERÊNCIA E, ACIMA DE TUDO, COM TOTAL LIBERDADE.

No verso do formulário você poderá acrescentar qualquer sugestão ou comentário que julge pertinente sobre o trabalho do **professor**.

Obrigado, mais uma vez, pelo seu tempo e colaboração.

xxxxx

Presidente da Comissão Própria de Avaliação – CPA/xxxx

## E-MAIL ENVIADO AOS PROFESSORES

Caro professor:

No período de 16 a 21 de maio, estaremos realizando a nossa tradicional pesquisa de opinião dos alunos sobre os professores.

Trata-se de uma pesquisa que julgamos importantíssima, uma vez que consideramos a opinião **séria, ética e coerente** dos alunos como um fator essencial para a melhoria da qualidade do ensino oferecido pelo xxxx.

Para o pleno êxito da pesquisa, é imprescindível, a exemplo dos semestres anteriores, que possamos contar com a inestimável colaboração dos Coordenadores e dos PROFESSORES da Instituição.

A pesquisa, a ser **realizada em sala de aula**, consta de um formulário contendo alguns itens para os quais o aluno deverá assinalar a opção de resposta que corresponde à sua opinião sobre a atividade do professor. Em cada horário o aluno emitirá a sua opinião sobre o trabalho desenvolvido pelo professor da respectiva disciplina.

Para garantir a confidencialidade das respostas, e oferecer ao aluno condições de resposta com mais tranquilidade e liberdade, solicitamos que o professor se ausente da sala de aula pelo tempo necessário (cerca de 15-20 min) e a aplicação da pesquisa seja conduzida por um aluno.

Os resultados das pesquisas serão entregues aos professores correspondentes e aos respectivos coordenadores para uma reflexão conjunta, **depois** que as pautas das disciplinas estiverem encerradas, isto é, estiverem lançadas no sistema as notas e as frequências dos alunos. Esta é uma forma que temos utilizado para reforçar junto ao aluno a mensagem de que o desempenho dele na disciplina será independente das opiniões emitidas.

Obrigado, mais uma vez, pelo seu tempo e colaboração.

xxxxxx

Presidente da Comissão Própria de Avaliação – CPA/xxxxx

**Anexo 4: Orientação para a aplicação dos questionários**

COMISSÃO PRÓPRIA DE AVALIAÇÃO – CPA NÚCLEO DE APOIO TÉCNICO E ASSESSORIA – NATA PESQUISA DE OPINIÃO DOS ALUNOS SOBRE OS PROFESSORES 1º SEMESTRE/2005
---

### **ORIENTAÇÃO PARA O PROFESSOR**

- 1) Escolha um aluno para ser o responsável pela aplicação dos formulários.
- 2) Entregue o envelope lacrado para o aluno escolhido.
- 3) Retire-se da sala para que os alunos possam responder à pesquisa.

### **ORIENTAÇÃO PARA O ALUNO RESPONSÁVEL PELA APLICAÇÃO DOS FORMULÁRIOS**

- 1) Abra o envelope.
- 2) Retire os formulários do envelope.
- 3) Distribua um formulário para cada aluno presente.
- 4) Recolha os formulários preenchidos.
- 5) Guarde os formulários (preenchidos e em branco) no envelope.
- 6) Devolva, no setor de pautas, o envelope com os formulários.