



UNIVERSIDADE FEDERAL DA PARAÍBA

IDENTIFICAÇÃO E AVALIAÇÃO DE VARIÁVEIS QUE
CAUSAM QUALIDADE NA EDUCAÇÃO DAS ESCOLAS DA
REDE MUNICIPAL DA CIDADE DE JOÃO PESSOA PARA O
ANO DE 2005

Alysson Lívio Vasconcelos Guedes

João Pessoa
2008

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

IDENTIFICAÇÃO E AVALIAÇÃO DE VARIÁVEIS QUE
CAUSAM QUALIDADE NA EDUCAÇÃO DAS ESCOLAS DA
REDE MUNICIPAL DA CIDADE DE JOÃO PESSOA PARA O
ANO DE 2005

Alysson Lívio Vasconcelos Guedes

Monografia apresentada ao
curso de Bacharelado em
Estatística da Universidade
Federal da Paraíba, em
cumprimento às exigências
para obtenção do grau de
Bacharel.

Orientador: Professor Dr. Ulisses Umbelino dos Anjos
Departamento de Estatística - UFPB

João Pessoa
2008

Alysson Lívio Vasconcelos Guedes

Bacharel em Estatística

IDENTIFICAÇÃO E AVALIAÇÃO DE VARIÁVEIS QUE
CAUSAM QUALIDADE NA EDUCAÇÃO DAS ESCOLAS DA
REDE MUNICIPAL DA CIDADE DE JOÃO PESSOA PARA O
ANO DE 2005

Monografia apresentada ao
curso de Bacharelado em
Estatística da Universidade
Federal da Paraíba, em
cumprimento às exigências
para obtenção do grau de
Bacharel.

Aprovada em 02 de maio de 2008, pela banca examinadora

Prof. Dr. Ulisses Umbelino dos Anjos

Prof. MSc. Jozemar Pereira dos Santos

AGRADECIMENTO

- Agradeço a Deus pai e a Santa a Maria por todas as bênçãos que recebi em toda a minha vida;
- Ao meu amado pai e irmãos, por serem fonte de amor e apoio em toda minha vida; e a minha mãe, que como exemplo de educadora possui a maior dentre todas as qualidades: o amor para com os alunos;
- Ao meu professor e amigo Ulisses Umbelino dos Anjos pela orientação neste trabalho;
- Ao professor Jozemar Pereira dos Santos pelos ensinamentos e por aceitar o convite a participar da banca examinadora;
- Ao meu primeiro professor de estatística João Agnaldo do Nascimento pelos seus ensinamentos e por aceitar o convite a participar da banca examinadora;
- A todos os professores e funcionários do Departamento de Estatística da UFPB pelos ensinamentos, amizade e contribuições para minha formação como Bacharel em Estatística;
- Aos meus colegas restantes da minha primeira turma: Carlos Sérgio, Joseilme “Jr.” Fernandes, Fábio Alencar, Maxwell “Max” Diniz, Thiago Cavalcante; e desistentes: Elaine Cristina, Ariadne “Adinha” Taís e Cleonice “Cléo” Aline;

- Aos meus amados amigos e companheiros de curso Renata Grigório, Juliana Ferreira, Juliana Kelly, André “Fica Tranquilo” Virginio, Adeilda “Dedinha” Fernandes e especialmente Lerivan “Lê” Ferreira da Silva;
- Aos demais colegas de curso, que se tornaram grandes amigos e futuros companheiros de trabalho.
- Aos meus amigos de toda a vida Gedvan “Ged” Dias, Luciana “Lú” Vital, Thiago “Gov” Gouveia e Nadja “Japinha” Cristina;

"Seja você quem for, seja qual for a posição social que você tenha na vida, a mais alta ou a mais baixa, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e com muita fé em Deus, que um dia você chega lá. De alguma maneira você chega lá."

Ayrton Senna

Lista de Quadros

QUADRO 1 - Codificação das variáveis do conjunto de dados.....
QUADRO 2 - Relação entre variáveis dependentes e independentes do exemplo 1.....
QUADRO 3 - Tradução do exemplo 1 para equações estruturais
QUADRO 4 - Descrição dos componentes do modelo estrutural
QUADRO 5 - Descrição dos componentes do modelo de mensuração.....
QUADRO 6 - Componentes adicionais do LISREL.....
QUADRO 7 - Subscritos de tamanho das matrizes.....
QUADRO 8 - Modelo LISREL do exemplo 2.....
QUADRO 9 - Resumo dos efeitos diretos, indiretos e totais.....
QUADRO 10 - Relação das variáveis dependentes e independentes do modelo.....
QUADRO 11 - Estimativas para as covariâncias do modelo - anos iniciais.....
QUADRO 12 - Estimativas para os coeficientes de regressão do modelo - anos iniciais.....
QUADRO 13 - Características do modelo final e medidas de ajuste - anos iniciais.....
QUADRO 14 - Teste para amostras independentes.....
QUADRO 15 - Efeitos padronizados - anos iniciais.....
QUADRO 16 - Estimativas para as covariâncias do modelo - anos finais.....
QUADRO 17 - Estimativas para os coeficientes de regressão do modelo - anos finais.....

QUADRO 18 - Características do modelo final e medidas de ajuste - anos finais
QUADRO 19 - Coeficiente de variação dos conjuntos de dados.....
QUADRO 20 - Efeitos padronizados - anos finais.....
QUADRO 21 - Matriz de correlação das variáveis - anos iniciais.....
QUADRO 22 - Matriz de correlação das variáveis - anos finais.....

Lista de tabelas

TABELA 1 – Saeb 1997: Proficiências médias e desvio Padrão.....

TABELA 2 – Limite superior e inferior das proficiências.....

TABELA 3 - Medidas descritivas das variáveis da educação básica - anos iniciais.....

TABELA 4 - Medidas descritivas das variáveis da educação básica - anos finais - ano 2005....

TABELA 5 - Distribuição de frequência média horas/aula diária - 4º série.....

TABELA 6 - Distribuição de frequência média horas/aula diária - 8º série.....

TABELA 7 - Distribuição de frequência de porcentagem de professores com curso superior - anos finais.....

Lista de figuras

FIGURA 1 - Diagrama de caminhos Exemplo

FIGURA 2 - Explicação das relações usadas no diagrama de caminhos

FIGURA 3 - Exemplo de modelo não-recursivo

FIGURA 4 - Relação pertencente ao diagrama de caminhos do exemplo 1

FIGURA 5 - Calculando os coeficientes estruturais para análise de caminhos

FIGURA 6 - Diagrama de caminhos exemplo 2

FIGURA 7 - Diagrama de caminhos obtido usando Modelagem de equações simultâneas

FIGURA 8 - Diagrama de caminhos do modelo sugerido

FIGURA 9 - Modelo final de equações simultâneas estimativas padronizadas Anos iniciais

FIGURA 10 - Modelo final de equações simultâneas estimativas não padronizadas Anos iniciais

FIGURA 11 - Histograma de Porcentagens de docentes com curso superior

FIGURA 12 - Boxplot Taxa de distorção idade-série e Taxa de Aprovação por Porcentagem de professores com curso superior

FIGURA 13 - Histograma alunos/turma anos iniciais

FIGURA 14 - Modelo final de equações simultâneas estimativas padronizadas Anos finais

FIGURA 15 - Modelo final de equações simultâneas estimativas não padronizadas Anos finais

Lista de siglas

AGFI - Adjusted goodness of fit index - Índice ajustado de qualidade

DETIIC - Diretoria de Tecnologia da Informação e Comunicação

FUNDEF - Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério

GFI - Goodness of fit index - Índice ajustado de qualidade

GL - Graus de liberdade

IDEB - Índice de Desenvolvimento da Educação Básica

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

LDB - Lei de Diretrizes e Bases

LISREL - Linear Structural Relationships

MEC - Ministério da Educação

PNE - Plano Nacional de Educação

PNUD - Programa das Nações Unidas para o Desenvolvimento

RMSA - Root mean square error of approximation - Raiz do erro quadrático médio da aproximação

Saeb - Sistema Nacional de Avaliação da Educação Básica

SEB - Secretaria de Educação Básica

SEM - Structural Equation Models - Modelagem de equações estruturais

UNICEF - The United Nations Children's Fund - Fundo das Nações Unidas para a Infância

Sumário

- 1 Introdução
- 2 Referencial Teórico
- 3 Objetivo
 - 3.1 Geral
 - 3.2 Específicos
- 4 Material e fonte dos dados
 - 4.1 Coleta
 - 4.2 Descrição das Variáveis.
 - 4.2.1 Indicadores de oferta
 - 4.2.1.1 Média de horas-aula diária.
 - 4.2.1.2 Porcentagem de docentes com curso superior.
 - 4.2.1.3 Número de alunos por turma.
 - 4.2.4 Indicadores de eficiência e rendimento
 - 4.2.4.1 Taxa de distorção idade-série
 - 4.2.4.2 Taxa de Aprovação
 - 4.2.5 Notas do Prova Brasil ou Saeb
 - 4.2.6 IDEB
 - 4.3 Organização dos dados
- 5 Metodologia
 - 5.1 Modelagem de Equações Estruturais
 - 5.1.1 Entendendo o SEM
 - 5.1.2 A Notação LISREL
 - 5.1.2.1 Estimando os parâmetros
 - 5.1.3 Modelagem de Equações Simultâneas
 - 5.1.4 Efeitos totais, diretos e indiretos

- 5.1.5 Medidas de bondade de ajuste
- 6 Resultados
 - 6.1 Avaliação descritiva
 - 6.1.1 Indicadores de oferta
 - 6.1.1.1 Média de horas-aula diária
 - 6.1.1.2 Porcentagem de docentes com curso superior
 - 6.1.1.3 Número de alunos por turma
 - 6.1.2 Indicadores de eficiência e rendimento
 - 6.1.2.1 Taxa de distorção idade-série
 - 6.1.2.2 Taxa aprovação
 - 6.1.3 Notas do Prova Brasil ou Saeb
 - 6.1.4 IDEB
 - 6.2 Modelagem de Equações Simultâneas
 - 6.3 Anos iniciais da educação básica
 - 6.4 Anos finais da educação básica
- 7 Conclusões
- 8 Referencias Bibliograficas

1 - Introdução

É de senso comum que educação de qualidade é fundamental para o desenvolvimento de um povo e um estado. No Brasil existe um esforço dos legisladores na elaboração de leis que visam assegurar um padrão mínimo de qualidade. A expressão *garantia de padrão de qualidade* [sobre a educação] aparece na Constituição Federal (BRASIL, CF/88, Art. 206) e na Lei de Diretrizes e Bases da Educação Nacional (LDB) (BRASIL, Lei nº 9.394, de 20 de dezembro de 1996, Art. 3º). No Plano Nacional de Educação (PNE) (BRASIL, MEC, 2001) é dito como Objetivo e Prioridade a *melhoria da qualidade do ensino em todos os níveis*; Sendo de obrigação da União prestar esse serviço de forma gratuita para toda população.

O que não existe é uma definição própria do que seria educação de qualidade. Nos últimos anos, principalmente após a promulgação do FUNDEF (BRASIL, Lei nº 9.424, de 24 de dezembro de 1996) que obriga o governo a determinar não mais o custo-aluno e sim o custo-aluno-qualidade, houve um aumento da discussão do que seria educação de qualidade. Levantando-se aí questões como: Quais as dimensões que medem a qualidade? Quais variáveis influenciam na qualidade?

Para compreender as relações entre a qualidade na educação e suas dimensões é necessário o suporte da ferramenta estatística modelagem de equações estruturais que permite o estudo simultâneo das diversas dimensões e variáveis associadas à qualidade.

Este estudo baseasse na necessidade de discriminar as variáveis que são responsáveis pela qualidade na educação no município de João Pessoa e subsidiar os gestores de políticas educacionais a adotar medidas mais efetivas na rede educacional municipal.

2 - Referencial Teórico

A legislação educacional do Brasil enquanto nação independente tem seu início na Constituição Imperial de 1824 (a qual continha um artigo sobre educação escolar primária gratuita) e prossegue até a Constituição Federal de 1988, considerando-se aí também as Constituições Estaduais, as Leis Orgânicas dos Municípios e toda a legislação ordinária, com ênfase especial na Lei de Diretrizes e Bases da Educação Nacional, nos diferentes momentos históricos em que elas ocorreram (MENEZES).

De acordo com Paro (1998), a educação, entendida como a apropriação do saber historicamente produzido é prática social que consiste na própria atualização cultural e histórica do homem. Na produção material de sua existência, as pessoas, juntamente com a construção de suas histórias, produzem conhecimento, valores, técnicas, atitudes, comportamentos, os quais configuram o saber historicamente produzido.

Essas proposições justificam a escolha das variáveis para a construção de um modelo de equações simultâneas que avalie as variáveis que causam qualidade no município de João Pessoa para o ano de 2005.

3 - Objetivo

Esta proposta de trabalho por sua natureza articula as áreas do conhecimento da Estatística e da Educação. Traçam-se, assim, os seguintes objetivos:

3.1 - Geral

Identificar e avaliar variáveis que causam qualidade na educação das escolas da rede municipal da cidade de João Pessoa para o ano de 2005.

3.2 - Específicos

- Construir um modelo teórico para as variáveis adquiridas

- Avaliar as relações presentes no modelo

4 - Material e fonte dos dados

A pesquisa foi realizada no município de João Pessoa para o ano de 2005 e tem como unidade amostral as escolas da rede municipal de ensino. A coleta foi realizada através de censo, tendo o município nesse período 90 escolas. As variáveis obtidas na pesquisa foram observadas para dois grupos distintos. O primeiro é para os anos iniciais da educação básica, série inicial a 4ª série, e o segundo é para os anos finais da educação básica, 5ª a 8ª série. Sendo retirados os casos com variáveis não observadas, exclusão *listwise*, restando então 73 escolas com informações sobre os anos iniciais da educação básica e 49 sobre os anos finais. A partir de 2006 o ensino básico passou a ter não mais 8 e sim 9 anos. Portanto atualmente os anos iniciais da educação básica seriam do 1º ao 5º ano e os anos finais do 6º ao 9º ano. Os estudos realizados são feitos para ambos os casos.

4.1 - Coleta

Devido ao pequeno tempo para a pesquisa, o número de variáveis coletadas foi menor que o esperado. Apesar do pouco tempo para coleta, tivemos ajuda dos funcionários da Secretaria de Educação do Município o que foi muito importante nessa etapa, devido à experiência profissional deles com os dados do censo escolar e as variáveis escolares. Os problemas encontrados foram principalmente devido à incompletude dos dados. Nem todas as variáveis podiam ser trabalhadas em todos os anos. Por isso o ano de 2005 foi escolhido para o estudo. Pois contava com a divulgação final do censo escolar 2006, usado para calcular os indicadores de rendimento entre outros. Outro problema era o uso de variáveis que podiam ser observadas por escola. O acesso a variáveis referentes ao município como um todo é mais fácil que o acesso aos valores por escola. Fora os dados fornecidos pela Secretaria de Educação, mais especificamente a Diretoria de Tecnologia da Informação e Comunicação (DTIC), as informações foram coletadas nos sites governamentais: <http://www.inep.gov.br/> e <http://portal.mec.gov.br/index.php>.

4.2 - Descrição das Variáveis.

4.2.1 - Indicadores de oferta

4.2.1.1 - Média de horas-aula diária.

Mede o tempo médio de horas-aula por dia em cada escola. A LDB (BRASIL, Lei nº 9.394, de 20 de dezembro de 1996, Art. 34º) diz que:

“A jornada escolar no ensino fundamental incluirá pelo menos quatro horas de trabalho efetivo em sala de aula, sendo progressivamente ampliado o período de permanência na escola”.

Método de calculo:

$$NHAD_{ks} = \frac{\sum_{h=0}^H h_{ks} \times M_{ksh}}{M_{ks}}$$

h_{ks} = número de horas-aula diária no nível de ensino k;

M_{ksh} = matricula no nível de ensino k, na série ou grupo de séries s, com h horas-aula diária;

M_{ks} = matriculas no nível de ensino k, na série ou grupo de séries s.

Onde:

k = creche, pré-escola, classe de alfabetização, ensino fundamental e ensino médio;

s = ensino fundamental (série, 1ª a 4ª, 5ª a 8ª, total), ensino médio (série e total).

4.2.1.2- Porcentagem de docentes com curso superior.

Percentual de docentes com formação superior em cada escola nas diferentes modalidades da educação básica. Por lei essa formação é obrigatória para os anos finais da educação e não obrigatória nos anos iniciais. Conforme a LDB (BRASIL, Lei nº 9.394, de 20 de dezembro de 1996, Art. 62º):

“A formação de docentes para atuar na educação básica far-se-á em nível superior, em curso de licenciatura, de graduação plena, em universidades e institutos superiores de educação, admitida, como

formação mínima para o exercício do magistério na educação infantil e nas quatro primeiras séries do ensino fundamental, a oferecida em nível médio, na modalidade Normal”.

Método de calculo:

$$PD_k^{ES} = \frac{D_k^{ES}}{D_k} \times 100$$

D_k^{ES} = número de docentes com formação superior, atuando no nível de ensino k;

D_k = número total de docentes atuando no nível de ensino k;

Onde:

k = creche, pré-escola, classe de alfabetização, ensino fundamental e ensino médio, educação especial e educação de jovens e adultos;

4.2.1.3 - Número de alunos por turma.

Número total de alunos matriculados na escola por número total de turmas. Não existindo valor legal definido, como vemos na LDB (BRASIL, Lei nº 9.394, de 20 de dezembro de 1996, Art. 25º):

“Será objetivo permanente das autoridades responsáveis alcançar relação adequada entre o número de alunos e o professor, a carga horária e as condições materiais do estabelecimento”.

Método do calculo:

$$NAT_{ks} = \frac{M_{ks}}{T_{ks}}$$

M_{ks} = matriculas no nível de ensino k, na série ou grupo de séries s;

T_{ks} = número de turmas no nível de ensino k, na série ou grupo de séries s.

4.2.4 - Indicadores de eficiência e rendimento

4.2.4.1 - Taxa de distorção idade-série

Expressa o percentual de alunos, em cada série, com idade superior à idade recomendada em cada escola.

Método de cálculo:

$$TDIS_{ks} = \frac{M_{ksi_sup}}{M_{ks}} \times 100$$

M_{ksi_sup} = número de matrículas na idade i_sup acima da recomendada para o nível de ensino k e na série ou grupo de séries s ;

M_{ks} = número total de matrículas no nível de ensino k na série ou grupo de séries s ;

s = Ensino fundamental (série, 1ª a 4ª, 5ª a 8ª, total), ensino médio (série, total).

4.2.4.2 - Taxa de Aprovação

É um índice de rendimento escolar e expressa o percentual de alunos aprovados.

$$TApov_s = \frac{Apov_s}{M_s - AF_s + AD_s + REC_s - DES_s} \times 100$$

$Apov_s$ = Número de aprovados na série ou grupo de séries s ;

M_s = Matrícula inicial na série ou grupo de séries s ;

AF_s = Afastados por abandono na série ou grupo de séries s ;

AD_s = Admitidos por transferência na série ou grupo de séries s ;

REC_s = Reclassificados para série ou grupo de séries s ;

DES_s = Alunos que saíram da série ou grupo de séries s , por reclassificação;

s = Ensino fundamental (série, 1ª a 4ª, 5ª a 8ª, total), ensino médio (série, total).

4.2.5 - Notas do Prova Brasil ou Saeb

O Sistema Nacional de Avaliação da Educação Básica – Saeb foi criado em 1995 sendo a primeira ferramenta de estudo, em nível nacional, sobre o sistema

educacional brasileiro. Ele avalia uma amostra representativa dos alunos regularmente matriculados nas 4ª e 8ª séries do ensino fundamental e 3º ano do ensino médio, de escolas públicas e privadas, localizadas em área urbana ou rural. Utiliza diferentes instrumentos de coleta de dados, sendo um deles os testes que têm por finalidade medir a habilidade de leitura em Língua Portuguesa e de resolução de problemas em Matemática dos alunos. Em 2005 foi criado o Prova Brasil que possui caráter censitário e é aplicado a todos os alunos da rede pública urbana de ensino, da 4ª e 8ª séries do ensino fundamental. Sendo o Saeb e o Prova Brasil complementares e realizados concomitantemente a cada dois anos, de forma que nenhum aluno será submetido às duas avaliações.

4.2.6 - IDEB

O IDEB, ou Índice de Desenvolvimento da Educação Básica, é um indicador de qualidade educacional que combina informações de desempenho em exames padronizados (Prova Brasil ou Saeb) obtido pelos estudantes ao final das etapas de ensino (4ª e 8ª séries do ensino fundamental e 3ª série do ensino médio) com informações sobre rendimento escolar.

Método de cálculo:

A forma geral do IDEB é dada por (1):

$$IDEB_{ji} = N_{ji}P_{ji} \quad 0 \leq j \leq 10 \ N; 0 \leq j \leq 1 \ P \text{ e } 0 \leq j \leq 10 \quad (1)$$

em que,

i = ano do exame (Saeb e Prova Brasil) e do Censo Escolar;

N_{ji} = média da proficiência em Língua Portuguesa e Matemática, padronizada para um indicador entre 0 e 10, dos alunos da unidade j , obtida em determinada edição do exame realizado ao final da etapa de ensino;

P_{ji} = indicador de rendimento baseado na taxa de aprovação da etapa de ensino dos alunos da unidade j ;

Em (1), a média de proficiência padronizada dos estudantes da unidade j , N_{ji} , é obtida a partir das proficiências médias em Língua Portuguesa e Matemática dos estudantes submetidos a determinada edição do exame realizado ao final da etapa educacional considerada (Prova Brasil ou Saeb). A proficiência média é padronizada

para estar entre zero e dez, de modo que $0 \leq IDEB \leq 10$. N_{ji} é obtida de acordo com (2).

$$N_{ji} = \frac{n_{ji}^{lp} + n_{ji}^{mat}}{2} \quad e \quad n_{ji}^{\alpha} = \frac{S_{ji}^{\alpha} - S_{inf}^{\alpha}}{S_{sup}^{\alpha} - S_{inf}^{\alpha}} * 10 \quad (2)$$

Em que,

n = proficiência na disciplina α , obtida pela unidade j , no ano i , padronizada para valores entre 0 e 10;

α = disciplina (Matemática ou Língua Portuguesa);

S_{ji}^{α} = proficiência média (em Língua Portuguesa ou Matemática), não padronizada, dos alunos da unidade j obtida no exame do ano i ;

S_{inf}^{α} = limite inferior da média de proficiência (Língua Portuguesa ou Matemática) do Saeb 1997;

S_{sup}^{α} = limite superior da média de proficiência (Língua Portuguesa ou Matemática) do Saeb 1997.

Para as unidades escolares (ou redes) que obtiverem $S_{ji}^{\alpha} < S_{inf}^{\alpha}$, a proficiência média é fixada em S_{inf}^{α} . Por sua vez, aquelas unidades que obtiverem $S_{ji}^{\alpha} > S_{sup}^{\alpha}$ têm o desempenho fixado em S_{sup}^{α} . A Tabela 1 apresenta a média e o desvio padrão das proficiências dos alunos da 4ª e da 8ª série do ensino fundamental e da 3ª série do ensino médio no Saeb de 1997. Posteriormente, a Tabela 2 traz os valores dos limites inferiores e superiores utilizados na padronização das proficiências médias em Língua Portuguesa e Matemática dos alunos da 4ª e da 8ª série do ensino fundamental e da 3ª série do ensino médio.

TABELA 1 – Saeb 1997: Proficiências médias e desvio Padrão

Série	Matemática		Língua Portuguesa	
	Média	Desvio Padrão	Média	Desvio Padrão
4ª do EF	190.8	44	186.5	46
8ª do EF	250.0	50	250.0	50
3ª do EM	288.7	59	283.9	56

Fonte: Saeb 1997 – Inep/MEC

A partir da média e desvio padrão das proficiências no Saeb 1997 (ano em que a escala do Saeb foi definida), calcularam-se, para cada etapa de ensino, considerando as diferentes disciplinas avaliadas no exame, os limites inferior e superior, de acordo com

$$S_{\text{inf}}^{\alpha} = \text{média}_{\alpha} - (3 * DP) \text{ e } S_{\text{sup}}^{\alpha} = \text{média}_{\alpha} + (3 * DP)$$

TABELA 2 – Limite superior e inferior das proficiências

Série	Matemática		Língua Portuguesa	
	S _{inf}	S _{sup}	S _{inf}	S _{sup}
4ª do EF	60	322	49	324
8ª do EF	100	400	100	400
3ª do EM	111	467	117	451

Fonte: Saeb 1997 – Inep/MEC

Esses limites, inferiores e superiores, apresentados na Tabela 2, são usados para calcular todos os IDEB's, ou seja, desde 1997, a partir do Saeb, para o Brasil (rede privada e pública; urbanas e rurais) e para os dados agregados por unidade da federação e, a partir da Prova Brasil de 2005, para municípios (rede municipal e estadual) e para as escolas.

O indicador de rendimento, P_j , é obtido conforme (3), onde a proporção de aprovados em cada uma das séries da etapa considerada, p^r , é calculada diretamente do Censo Escolar¹. Se p^r ($r = 1, 2, \dots, n$, em que n é o número de séries com taxa de aprovação positiva) é a taxa de aprovação da r -ésima série da etapa educacional considerada, então o tempo médio de duração da série é:

$$T_{ji} = \sum_{r=1}^n \frac{1}{p^r} = \frac{n}{P_{ji}} \quad (3)$$

Em (3), P_{ji} é a taxa média de aprovação na etapa educacional no ano i . Note-se que, na ausência de evasão durante a etapa e em equilíbrio estacionário, $\frac{n}{P_{ji}}$ dá o tempo médio para conclusão de uma etapa para os estudantes da unidade j (T_{ji}).

Se P é o inverso do tempo médio para conclusão de uma série, então, $P_{ji} = \frac{1}{T_{ji}}$.

Deste modo, temos que $IDEB_{ji} = \frac{N_{ji}}{T_{ji}}$, ou seja, o indicador fica sendo a pontuação no

exame padronizado ajustada pelo tempo médio (em anos) para conclusão de uma série naquela etapa de ensino.

1 Para o cálculo de T_{ji} , o tempo médio para a conclusão de uma série, foram utilizados os dados de aprovação de cada uma das séries das três etapas de ensino: 1ª e 2ª fase do ensino fundamental e ensino médio. Para a operacionalização do indicador, convencionou-se o critério de considerar como inexistentes as séries que apresentavam taxa de aprovação zero.

4.3 - Organização dos dados

Os dados foram organizados em planilha eletrônicas do Microsoft Excel 2007, referentes às 90 escolas da rede municipal de educação da cidade de João Pessoa no ano de 2005.

As variáveis selecionadas foram transferidas para um banco de dados nos pacotes SPSS 13.0, onde foram realizados os estudos descritivos, e para o AMOS 6, onde foram realizadas as análises dos dados através da técnica multivariada modelagem de equações simultâneas.

os dados foram codificados da seguinte forma:

QUADRO 1 - Codificação das variáveis do conjunto de dados

X1	Média horas/aula diária
X2	Porcentagem de docentes com curso superior
X3	Número de Alunos por turma
Y1	Taxa de Distorção Idade-Série
Y2	Taxa de Aprovação
Y3	Matemática
Y4	Língua Portuguesa
Y5	IDEB

5 - Metodologia

5.1 - Modelagem de Equações Estruturais

A modelagem de equações estruturais, também conhecida como SEM (Structural Equation Modeling), é uma técnica Multivariada que permite o uso de relações separadas para cada conjunto de variáveis dependentes. O SEM nos fornece uma técnica de estimação apropriada e mais eficiente quando se tem um conjunto de equações de regressão múltipla separadas, mas interdependentes, que devem ser estimadas simultaneamente. O SEM é caracterizado por dois componentes básicos: O modelo estrutural e o modelo de mensuração. O modelo estrutural é um conjunto de uma ou mais relações de dependência entre as variáveis latentes (construtos) do modelo; o modelo de mensuração especifica quais variáveis observadas (indicadores) que serão utilizadas como medida para cada variável latente, além de poder avaliar a confiabilidade desses indicadores para medir as variáveis latentes associadas. O SEM é resultante de uma evolução da modelagem de múltiplas equações desenvolvida principalmente nas ciências econométricas e originada dos princípios de mensuração da psicologia e sociologia. Em termos simples, o SEM combina aspectos de regressão múltipla com análise fatorial para estimar uma série de relações de dependência inter-relacionadas simultaneamente.

5.1.1 - Entendendo o SEM

O SEM permite a compreensão de relações subseqüentes de variáveis, ou seja, variáveis independentes, explicando variáveis dependentes e estas por sua vez, explicando outras variáveis dependentes. Gerando subseqüentes relações que a Regressão Múltipla não pode resolver. A maneira mais simples de formular o SEM é:

$$\begin{aligned} Y_1 &= X_{11} + X_{12} + \dots + X_{1n} \\ Y_2 &= X_{21} + X_{22} + \dots + X_{2n} \\ Y_m &= X_{m1} + X_{m2} + \dots + X_{mn} \end{aligned}$$

Estas relações do SEM podem ser aplicadas a **construtos**. Construtos são idéias ou conceitos que não podem ser medidos diretamente e por isso são definidos por conjuntos de variáveis. Logo conceitos como inteligência seriam medidos por variáveis

como QI, velocidade de cálculos e etc. Observe que não podemos medir o construto diretamente e sim por **indicadores**, portanto, o construto Inteligência é medido pelos indicadores QI, velocidade de cálculos e etc. As variáveis ou construtos que não podem ser medidos diretamente são chamados de **latentes**. Observe que não medimos diretamente o construto inteligência, logo ele é um construto latente. E as variáveis que podem ser observadas diretamente são chamadas de **manifestas**. Como podemos medir QI (existem vários testes para medir o QI de uma pessoa) e o tempo que ela demora a realizar certas contas, chamamos estas variáveis de manifestas.

Exemplo 1: Uma empresa deseja aumentar a retenção de empregados, para isso pesquisou sobre seus empregados e seus relacionamentos no trabalho. Para isso foram identificados 6 variáveis importantes para o estudo: *Atitude com os colegas de trabalho*, *Ambiente de trabalho*, *Nível de pagamento*, *Satisfação com o emprego*, *Compromisso organizacional* e *Probabilidade de mudança de empregado*. Verificam-se as seguintes relações entre essas variáveis: *Atitude com os colegas de trabalho* e *Ambiente de trabalho* explicam *Satisfação com o emprego*. *Nível de pagamento* e *Satisfação com o emprego* explicam *Compromisso organizacional* e por fim *Satisfação com o emprego* e *Compromisso organizacional* explicam *Probabilidade de mudança de empregado*. Observe o quadro 2.

QUADRO 2 - Relação entre variáveis dependentes e independentes do exemplo 1

Variável dependente	Variáveis independentes
<i>Satisfação com o emprego</i>	= <i>Atitude com os colegas de trabalho</i> + <i>Ambiente de trabalho</i>
<i>Compromisso organizacional</i>	= <i>Satisfação com o emprego</i> + <i>Nível de pagamento</i>
<i>Probabilidade de mudança de empregado</i>	= <i>Satisfação com o emprego</i> + <i>Compromisso organizacional</i>

Veja que *Satisfação com o emprego* é na primeira relação uma variável dependente e torna-se uma variável independente na segunda e terceira relação. O *Compromisso organizacional* que na segunda relação é uma variável dependente torna-se independente na terceira relação.

Veja que nessas relações consecutivas temos variáveis que explicam e outras que são explicadas. Chamamos as variáveis ou construtos que apenas explicam de **exógenos** e as denotamos de **X**. As variáveis ou construtos que são explicadas por pelo menos uma relação causal de chamamos de **endógenos** e denotamos por **Y**. Portanto *Atitude com os colegas de trabalho*, *Ambiente de trabalho*, *Nível de pagamento* são construtos exógenos, pois apenas explicam as outras variáveis e *Satisfação com o emprego*, *Compromisso organizacional*, *Probabilidade de mudança de empregado* são construtos endógenos, pois são explicados em pelo menos uma relação.

Estas relações podem ser descritas visualmente através do **Diagrama de Caminhos**:

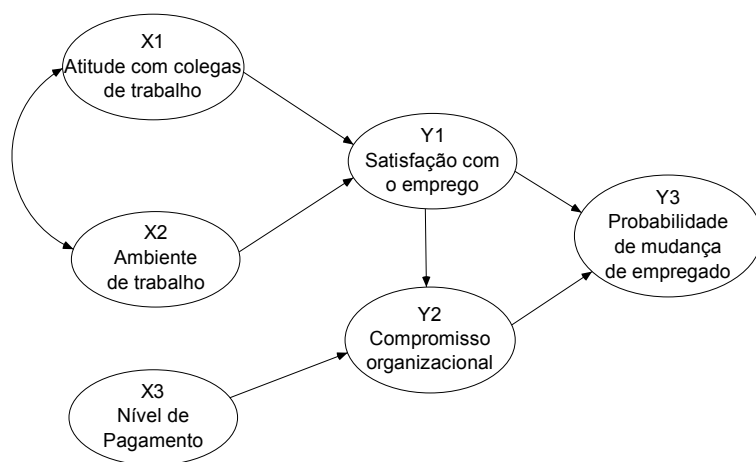


FIGURA 1 - Diagrama de caminhos Exemplo 1

O diagrama de caminhos é a representação gráfica do SEM. No diagrama as elipses representam as variáveis não observadas (latentes) ou os erros, as setas retilíneas indicam as relações causais que apontando da variável preditora à variável ou construto dependente. As setas curvas correspondem a correlações ou covariâncias entre variáveis, não indicando causalidade entre elas. Vemos diagramas também que apresentam quadrados ou retângulos, estes representam variáveis observadas. Veja a figura abaixo:

Símbolo

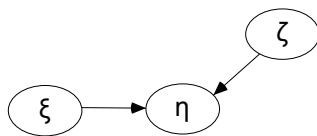
Explicação



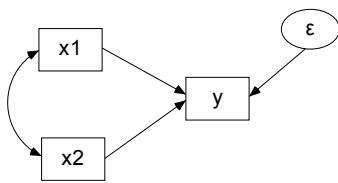
Variáveis manifestas (observadas)



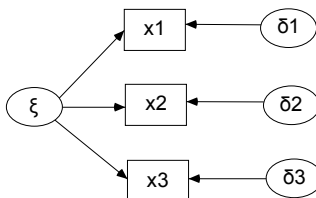
Variáveis latentes (Não observadas)



Variável latente causando outra variável latente com um erro



Duas variáveis manifestas associadas explicando outra manifesta com um erro



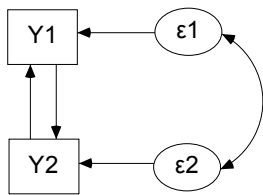
Variável latente causando variáveis manifestas com seus próprios erros

FIGURA 2 - Explicação das relações usadas no diagrama de caminhos

Os diagramas devem ser desenhados de forma que o fluxo causal seja da esquerda para a direita. Podemos ainda classificar o SEM de duas maneiras: **recursivo** e **não-recursivo**. Um modelo é recursivo quando a causalidade acontece em somente uma direção, assim as setas tem um único sentido. Em um modelo não-recursivo o fluxo causal pode ter mais de um sentido. A exemplo:

Símbolo

Explicação



Modelo não-recursivo entre duas variáveis manifestas com erros correlacionados

FIGURA 3 - Exemplo de modelo não-recursivo

O diagrama de caminhos é a base para a modelagem de equações simultâneas, método que emprega correlações bivariadas simples para estimar as relações em um sistema de equações estruturais. O método é baseado na especificação das relações em uma série de equações do tipo regressão (retratadas graficamente em diagrama de caminhos) que podem então ser estimados pela determinação da quantia de correlação atribuível a cada efeito em cada equação simultaneamente. Quando empregada com múltiplas relações entre construtos latentes e um modelo de mensuração, ela é então chamada de modelagem de equações estruturais (HAIR, 2005). Para ilustrar usaremos uma parte da relação presente na figura 1.

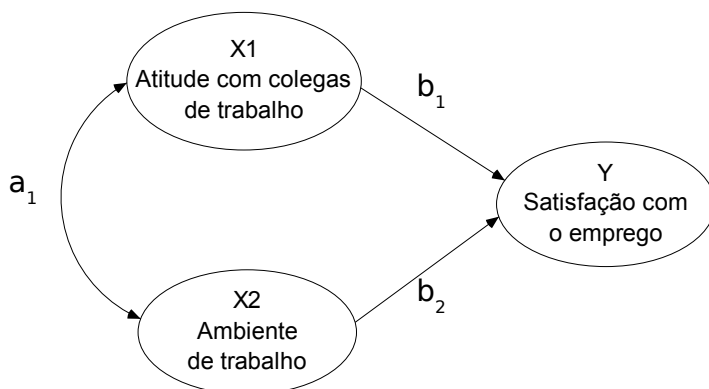


FIGURA 4 - Relação pertencente ao diagrama de caminhos do exemplo 1

Padronizamos os dados de forma que:

$$E(X_1) = E(X_2) = E(Y) = 0$$

$$V(X_1) = V(X_2) = V(Y) = 1$$

(4)

Sabemos que o modelo de regressão para este exemplo com os dados padronizados obedece à seguinte equação:

$$Y = b_1 X_1 + b_2 X_2 + \varepsilon \quad (5)$$

Substituindo na correlação simples temos:

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{\mathbf{E}(X \cdot Y) - \mathbf{E}(X) \cdot \mathbf{E}(Y)}{\sqrt{V(X) \cdot V(Y)}}$$

$$\begin{cases} Cor(X, Y) = \mathbf{E}(X \cdot Y) \\ Cor(X_1, X_2) = \mathbf{E}(X_1 \cdot X_2) \end{cases} \quad (6)$$

Substituindo (5) em $\mathbf{E}(X_1 Y)$:

$$\begin{aligned} \mathbf{E}(X_1 Y) &= \mathbf{E}(X_1 \cdot (b_1 X_1 + b_2 X_2)) \\ &= b_1 \cdot \mathbf{E}(X_1^2) + b_2 \cdot \mathbf{E}(X_1 \cdot X_2) \end{aligned} \quad (7)$$

De (4) temos

$$\begin{aligned} V(X) &= \mathbf{E}(X^2) - \mathbf{E}(X)^2 \\ 1 &= \mathbf{E}(X^2) - 0 \\ \mathbf{E}(X^2) &= 1 \end{aligned} \quad (8)$$

Utilizando (6) e (8) em (7) e chamando a correlação de X_1 com X_2 de a_1 temos:

$$\begin{aligned} \mathbf{E}(X_1 Y) &= b_1 + b_2 \cdot \mathbf{E}(X_1 \cdot X_2) \\ &= b_1 + b_2 \cdot Cor(X_1 \cdot X_2) \\ &= b_1 + b_2 \cdot a_1 \end{aligned} \quad (9)$$

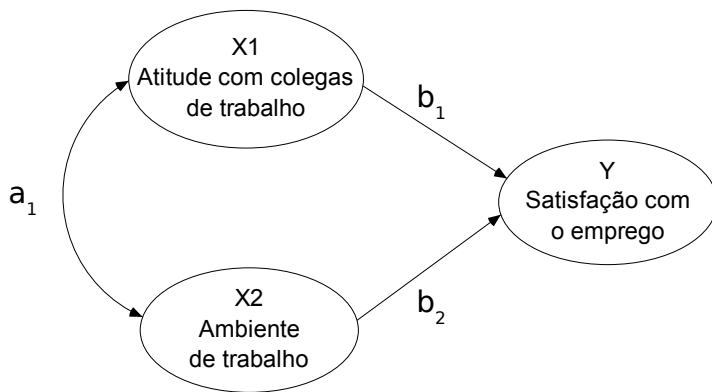
De maneira similar temos

$$\begin{aligned} \mathbf{E}(X_2 Y) &= \mathbf{E}(X_2 \cdot (b_1 X_1 + b_2 X_2)) \\ &= b_2 \cdot \mathbf{E}(X_2^2) + b_1 \cdot \mathbf{E}(X_1 \cdot X_2) \\ &= b_2 + b_1 \cdot Cor(X_1 \cdot X_2) \\ &= b_2 + b_1 \cdot a_1 \end{aligned} \quad (10)$$

Por fim temos de (9) e (10)

$$\begin{cases} \text{Cor}(X_1 X_2) = a_1 \\ \text{Cor}(X_1 Y) = b_1 + b_2 \cdot a_1 \\ \text{Cor}(X_2 Y) = b_2 + b_1 \cdot a_1 \end{cases} \quad (11)$$

Logo vemos que a correlação simples entre duas variáveis pode ser representada como a soma dos caminhos compostos conectando esses pontos.



Matriz de correlação

$$\begin{matrix} & Y & X_1 & X_2 \\ Y & 1 & & \\ X_1 & 0,6 & 1 & \\ X_2 & 0,7 & 0,5 & 1 \end{matrix}$$

$$\begin{cases} \text{Cor}(X_1 X_2) = a_1 \\ \text{Cor}(X_1 Y) = b_1 + b_2 \cdot a_1 \\ \text{Cor}(X_2 Y) = b_2 + b_1 \cdot a_1 \end{cases} \quad \begin{cases} 0,5 = a_1 \\ 0,6 = b_1 + b_2 \cdot 0,5 \\ 0,7 = b_2 + b_1 \cdot 0,5 \end{cases}$$

FIGURA 5 - Calculando os coeficientes estruturais para análise de caminhos

Veja que podemos aplicar $b_2 = 0,53$ para todas as relações do diagrama de caminhos, pois só foi necessário apenas a matriz de covariâncias ou de correlação. Podemos agora definir de forma específica o modelo de mensuração do SEM para o diagrama de caminhos da figura 1 no quadro 2.

QUADRO 3 - Tradução do exemplo 1 para equações estruturais

Variável Endógena	=	Variáveis Exógenas	+	Variáveis endógenas	+	Erro
Y1	=	$b_1 X_1 + b_2 X_2$	+		+	ϵ_1
Y2	=	$b_3 X_3$	+	$b_4 Y_1$	+	ϵ_2

Y3	=	+	b ₅ Y ₁ +b ₆ Y ₂	+	ε ₃
----	---	---	--	---	----------------

5.1.2 - A Notação LISREL

Existem muitas notações para descrever o Modelo de Equações Estruturais. Dentre estas notações a mais conhecida é a LISREL. A grande vantagem de usar esses sistemas de notação é a ajuda para descobrir os coeficientes estruturais quando usamos variáveis latentes. A notação LISREL é formada basicamente por um conjunto de 8 matrizes (sem contar o erro). A equação para o modelo estrutural que descreve a relação entre os construtos endógenos e exógenos é:

$$\eta = \Gamma\xi + B\eta + \zeta$$

QUADRO 4 - Descrição dos componentes do modelo estrutural

Matriz	Nome	Descrição	Dimensões	Elementos
η	Eta	Construtos endógenos	n x 1	
Γ	Gama	Relação dos construtos endógenos com exógenos	n x m	γ _{nm}
ξ	Ksi	Construtos exógenos	m x 1	
B	Beta	Relações de construtos endógenos com endógenos	n x n	β _{nn}
ζ	Zeta	Erro	n x 1	

As equações para modelo de mensuração são os seguintes:

$$X = \Lambda_x \xi + \delta$$

$$Y = \Lambda_y \eta + \varepsilon$$

QUADRO 5 - Descrição dos componentes do modelo de mensuração

Matriz	Nome	Descrição	Dimensões	Elemento
X		Indicador exógeno	p x 1	
Y		Indicador endógeno	q x m	
Λ _x	Lambida-X	Cargas dos indicadores exógenos	p x m	λ ^x _{pm}
Λ _y	Lambida-Y	Cargas dos indicadores endógenos	q x n	λ ^y _{qn}
δ	Delta	Erro	p x 1	
ε	Epsilon	Erro	q x 1	

O modelo LISREL ainda pede alguns elementos

QUADRO 6 - Componentes adicionais do LISREL

Matriz	Nome	Descrição	Dimensões	Elementos
Φ	Phi	Correlação entre os construtos exógenos	$m \times m$	ϕ_{mm}
Ψ	Psi	Correlação entre os construtos endógenos	$n \times n$	ψ_{nn}
Θ_{δ}	Theta-Delta	Matriz de correlação de erro de previsão para indicadores de construtos exógenos	$p \times p$	θ_{pp}^{δ}
Θ_{ϵ}	Theta-Epsilon	Matriz de correlação de erro de previsão para indicadores de construtos endógenos	$q \times q$	θ_{qq}^{ϵ}

QUADRO 7 - Subscritos de tamanho das matrizes

Subscrito	Descrição
m	Número de construtos exógenos
n	Número de construtos endógenos
p	Número de indicadores de construtos exógenos
q	Número de indicadores de construtos endógenos

5.1.2.1 - Estimando os parâmetros

Para a estimação dos parâmetros assumimos as seguintes pressuposições:

$$\begin{aligned}
 \mathbf{E}(\zeta) &= 0 & \text{Cov}(\zeta) &= \Psi \\
 \mathbf{E}(\epsilon) &= 0 & \text{Cov}(\epsilon) &= \Theta_{\epsilon} \\
 \mathbf{E}(\delta) &= 0 & \text{Cov}(\delta) &= \Theta_{\delta} \\
 \mathbf{E}(\xi) &= 0 & \text{Cov}(\xi) &= \Phi
 \end{aligned}$$

Onde ζ , ϵ , e δ são mutuamente não correlacionados; ζ é não correlacionado com ξ ; ϵ é não correlacionado com η ; δ é não correlacionado com ξ ; B possui apenas zeros na diagonal principal; (I - B) é não singular (JOHNSON, 1982).

As pressuposições do modelo implicam em certa estrutura de covariância.

$$\Sigma(\theta)_{(p+q) \times (p+q)} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \text{Cov} \begin{pmatrix} Y \\ X \end{pmatrix} = \begin{bmatrix} \text{Cov}(Y) & \text{Cov}(YX) \\ \text{Cov}(XY) & \text{Cov}(X) \end{bmatrix}$$

Onde

$$\begin{aligned} \eta &= \Gamma \xi + B \eta + \zeta \\ \eta - B \eta &= \Gamma \xi + \zeta \\ (\mathbf{I} - B) \eta &= \Gamma \xi + \zeta \\ \eta &= (\mathbf{I} - B)^{-1} \Gamma \xi + (\mathbf{I} - B)^{-1} \zeta \end{aligned}$$

$$\begin{aligned} \text{Cov}(\eta) &= \mathbf{E}(\eta \eta') = \mathbf{E}\{[(\mathbf{I} - B)^{-1} \Gamma \xi + (\mathbf{I} - B)^{-1} \zeta][(\mathbf{I} - B)^{-1} \Gamma \xi + (\mathbf{I} - B)^{-1} \zeta]'\} \\ &= \mathbf{E}\{[(\mathbf{I} - B)^{-1} \Gamma \xi + (\mathbf{I} - B)^{-1} \zeta][\xi' \Gamma (\mathbf{I} - B)^{-1} + \zeta' (\mathbf{I} - B)^{-1}]\} \\ &= \mathbf{E}\{[(\mathbf{I} - B)^{-1} \Gamma \xi \xi' \Gamma (\mathbf{I} - B)^{-1} + (\mathbf{I} - B)^{-1} \Gamma \xi \zeta' (\mathbf{I} - B)^{-1} \\ &\quad + (\mathbf{I} - B)^{-1} \zeta \xi' \Gamma (\mathbf{I} - B)^{-1} + (\mathbf{I} - B)^{-1} \zeta \zeta' (\mathbf{I} - B)^{-1}]\} \\ &= (\mathbf{I} - B)^{-1} \Gamma \mathbf{E}(\xi \xi') \Gamma (\mathbf{I} - B)^{-1} + (\mathbf{I} - B)^{-1} \Gamma \mathbf{E}(\xi \zeta') (\mathbf{I} - B)^{-1} \\ &\quad + (\mathbf{I} - B)^{-1} \mathbf{E}(\zeta \xi') \Gamma (\mathbf{I} - B)^{-1} + (\mathbf{I} - B)^{-1} \mathbf{E}(\zeta \zeta') (\mathbf{I} - B)^{-1} \\ &= (\mathbf{I} - B)^{-1} \Gamma \text{Cov}(\xi) \Gamma (\mathbf{I} - B)^{-1} + (\mathbf{I} - B)^{-1} \mathbf{E}(\zeta \zeta') (\mathbf{I} - B)^{-1} \\ &= (\mathbf{I} - B)^{-1} [\Gamma \Phi \Gamma + \Psi] (\mathbf{I} - B)^{-1} \end{aligned}$$

$$\begin{aligned} \text{Cov}(Y) &= \mathbf{E}(Y Y') = \mathbf{E}[(\Lambda_y \eta + \varepsilon)(\Lambda_y \eta + \varepsilon)'] \\ &= \mathbf{E}[(\Lambda_y \eta + \varepsilon)(\eta' \Lambda_y' + \varepsilon')] \\ &= \mathbf{E}[(\Lambda_y \eta \eta' \Lambda_y' + \Lambda_y \eta \varepsilon' + \varepsilon \eta' \Lambda_y' + \varepsilon \varepsilon')] \\ &= \Lambda_y \mathbf{E}(\eta \eta') \Lambda_y' + \Lambda_y \mathbf{E}(\eta \varepsilon') + \mathbf{E}(\varepsilon \eta') \Lambda_y' + \mathbf{E}(\varepsilon \varepsilon') \\ &= \Lambda_y \text{Cov}(\eta) \Lambda_y' + \text{Cov}(\varepsilon) \\ &= \Lambda_y [(\mathbf{I} - B)^{-1} (\Gamma \Phi \Gamma + \Psi) (\mathbf{I} - B)^{-1}] \Lambda_y' + \Theta_\varepsilon \end{aligned}$$

$$\begin{aligned} \text{Cov}(X) &= \mathbf{E}(X X') = \mathbf{E}[(\Lambda_x \xi + \delta)(\Lambda_x \xi + \delta)'] \\ &= \mathbf{E}[(\Lambda_x \xi + \delta)(\Lambda_x \xi + \delta)'] \\ &= \mathbf{E}[(\Lambda_x \xi + \delta)(\xi' \Lambda_x' + \delta')] \\ &= \mathbf{E}[(\Lambda_x \xi \xi' \Lambda_x' + \delta \delta')] \\ &= \Lambda_x \mathbf{E}(\xi \xi') \Lambda_x' + \mathbf{E}(\delta \delta') \\ &= \Lambda_x \Phi \Lambda_x' + \Theta_\delta \end{aligned}$$

$$\begin{aligned}
Cov(XY) &= Cov(YX)' \\
&= (\Lambda_y (\mathbf{I} - B)^{-1} \Gamma \Phi \Lambda_x')' \\
&= \Lambda_x' \Phi \Gamma (\mathbf{I} - B)^{-1} \Lambda_y'
\end{aligned}$$

$$\begin{aligned}
Cov(YX) &= E(YX') = E[(\Lambda_y \eta + \varepsilon)(\Lambda_x \xi + \delta)'] \\
&= E[(\Lambda_y \eta + \varepsilon)(\xi' \Lambda_x' + \delta')] \\
&= E(\Lambda_y \eta \xi' \Lambda_x' + \Lambda_y \eta \delta' + \varepsilon \xi' \Lambda_x' + \varepsilon \delta') \\
&= E\{\Lambda_y [(\mathbf{I} - B)^{-1} \Gamma \xi + (\mathbf{I} - B)^{-1} \zeta] \xi' \Lambda_x' + \Lambda_y [(\mathbf{I} - B)^{-1} \Gamma \xi + (\mathbf{I} - B)^{-1} \zeta] \delta' + \varepsilon \xi' \Lambda_x' + \varepsilon \delta'\} \\
&= E[\Lambda_y (\mathbf{I} - B)^{-1} \Gamma \xi \xi' \Lambda_x' + \Lambda_y (\mathbf{I} - B)^{-1} \zeta \zeta' \Lambda_x' \\
&\quad + \Lambda_y (\mathbf{I} - B)^{-1} \Gamma \xi \delta' + \Lambda_y (\mathbf{I} - B)^{-1} \zeta \delta' + \varepsilon \xi' \Lambda_x' + \varepsilon \delta'] \\
&= \Lambda_y (\mathbf{I} - B)^{-1} \Gamma E(\xi \xi') \Lambda_x' + \Lambda_y (\mathbf{I} - B)^{-1} E(\zeta \zeta') \Lambda_x' \\
&\quad + \Lambda_y (\mathbf{I} - B)^{-1} \Gamma E(\xi \delta') + \Lambda_y (\mathbf{I} - B)^{-1} E(\zeta \delta') + E(\varepsilon \xi') \Lambda_x' + E(\varepsilon \delta') \\
&= \Lambda_y (\mathbf{I} - B)^{-1} \Gamma Cov(\xi) \Lambda_x' \\
&= \Lambda_y (\mathbf{I} - B)^{-1} \Gamma \Phi \Lambda_x'
\end{aligned}$$

Destes resultados obtemos

$$\Sigma(\theta)_{(p+q) \times (p+q)} = \begin{bmatrix} \Lambda_y [(\mathbf{I} - B)^{-1} (\Gamma \Phi \Gamma' + \Psi) (\mathbf{I} - B)^{-1}] \Lambda_y' + \Theta_\varepsilon & \Lambda_y (\mathbf{I} - B)^{-1} \Gamma \Phi \Lambda_x' \\ \Lambda_x' \Phi \Gamma (\mathbf{I} - B)^{-1} \Lambda_y' & \Lambda_x' \Phi \Lambda_x' + \Theta_\delta \end{bmatrix}$$

A matriz de variância amostral é usada para estimar os parâmetros do modelo.

$$S_{(p+q) \times (p+q)} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \hat{\Sigma}$$

Para que o modelo não forme um sistema indeterminado, com mais variáveis indefinidas que conhecidas. Definimos um dos indicadores de cada construto em 1 tornando assim o sistema possível de ser resolvido.

Exemplo 2: Para um modelo com um construto exógeno e outro endógeno, com dois indicadores cada, ou seja, m=1, n=1, p=2, q=2 e B=0. Tem-se que:

QUADRO 8 - Modelo LISREL do exemplo 2

Modelo estrutural	$\eta_1 = \gamma_1 \xi_1 + \zeta_1$
Modelos de mensuração	$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} \eta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$ $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix} \xi + \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$

Adicionalmente temos as seguintes matrizes

$$\Psi = [\psi] \quad \Theta_\varepsilon = \begin{bmatrix} \theta_1 & \\ & \theta_2 \end{bmatrix} \quad \Theta_\delta = \begin{bmatrix} \theta_3 & \\ & \theta_4 \end{bmatrix}$$

O diagrama de caminhos para esta situação é apresentada na figura 6.

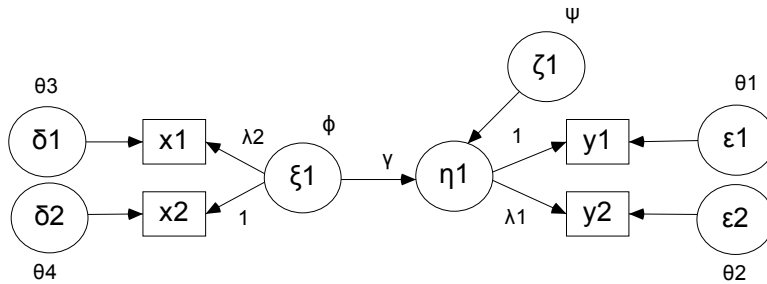


FIGURA 6 - Diagrama de caminhos exemplo 2

Assim, para estimar os parâmetros:

$$\Sigma(\theta) = \begin{bmatrix} \Lambda_y(\Gamma\Phi\Gamma' + \Psi)\Lambda_y' + \Theta_\varepsilon & \Lambda_y\Gamma\Phi\Lambda_x' \\ \Lambda_x\Phi\Gamma'\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \Theta_\delta \end{bmatrix}$$

Logo

$$\begin{aligned} Cov(Y) &= \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} (\gamma\phi\gamma + \psi) \begin{bmatrix} 1 & \lambda_1 \end{bmatrix} + \begin{bmatrix} \theta_1 & \\ & \theta_2 \end{bmatrix} \\ &= \begin{bmatrix} \gamma^2\phi + \psi \\ \lambda_1(\gamma^2\phi + \psi) \end{bmatrix} \begin{bmatrix} 1 & \lambda_1 \end{bmatrix} + \begin{bmatrix} \theta_1 & \\ & \theta_2 \end{bmatrix} \\ &= \begin{bmatrix} \gamma^2\phi + \psi & \lambda_1(\gamma^2\phi + \psi) \\ \lambda_1(\gamma^2\phi + \psi) & \lambda_1^2(\gamma^2\phi + \psi) \end{bmatrix} + \begin{bmatrix} \theta_1 & \\ & \theta_2 \end{bmatrix} \\ &= \begin{bmatrix} \gamma^2\phi + \psi + \theta_1 & \lambda_1(\gamma^2\phi + \psi) \\ \lambda_1(\gamma^2\phi + \psi) & \lambda_1^2(\gamma^2\phi + \psi) + \theta_2 \end{bmatrix} \end{aligned}$$

$$\begin{aligned}
Cov(X) &= \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix} \phi \begin{bmatrix} \lambda_2 & 1 \end{bmatrix} + \begin{bmatrix} \theta_3 & \\ & \theta_4 \end{bmatrix} \\
&= \begin{bmatrix} \phi \lambda_2 \\ \phi \end{bmatrix} \begin{bmatrix} \lambda_2 & 1 \end{bmatrix} + \begin{bmatrix} \theta_3 & \\ & \theta_4 \end{bmatrix} \\
&= \begin{bmatrix} \phi \lambda_2^2 & \phi \lambda_2 \\ \phi \lambda_2 & \phi \end{bmatrix} + \begin{bmatrix} \theta_3 & \\ & \theta_4 \end{bmatrix} \\
&= \begin{bmatrix} \phi \lambda_2^2 + \theta_3 & \phi \lambda_2 \\ \phi \lambda_2 & \phi + \theta_4 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
Cov(YX) &= \Lambda_y \Gamma \Phi \Lambda'_x \\
&= \begin{bmatrix} 1 \\ \lambda_4 \end{bmatrix} \gamma \phi \begin{bmatrix} \lambda_2 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \gamma \phi \\ \lambda_4 \gamma \phi \end{bmatrix} \begin{bmatrix} \lambda_2 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \lambda_2 \gamma \phi & \gamma \phi \\ \lambda_4 \lambda_2 \gamma \phi & \lambda_4 \gamma \phi \end{bmatrix}
\end{aligned}$$

$$Cov(XY) = \begin{bmatrix} \lambda_2 \gamma \phi & \lambda_4 \lambda_2 \gamma \phi \\ \gamma \phi & \lambda_4 \gamma \phi \end{bmatrix}$$

Das variáveis observadas retiramos a matriz de variância amostral

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} 14,3 & -27,6 & 6,4 & 3,2 \\ -27,6 & 55,4 & -12,8 & -6,4 \\ 6,4 & -12,8 & 3,7 & 1,6 \\ 3,2 & -6,4 & 1,6 & 1,1 \end{bmatrix}$$

Fazemos $\hat{\Sigma} = S$ para obter as seguintes equações

- ~~(i) \dots~~
- ~~(ii) \dots~~
- ~~(iii) \dots~~
- ~~(iv) \dots~~
- ~~(v) \dots~~
- ~~(vi) \dots~~
- ~~(vii) \dots~~
- ~~(viii) \dots~~
- ~~(ix) \dots~~
- ~~(x) \dots~~

De (iv) e (v) temos $\hat{\lambda}_2 = 2$. De (ix) temos $\hat{\phi} = 0,8$. De (v) temos $\hat{\gamma} = 4$.

Resolvendo as demais relações encontramos os seguintes resultados;

$$\hat{\gamma} = 4 \quad \hat{\Lambda}_y = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \hat{\Theta}_\varepsilon = \begin{bmatrix} 0,5 & \\ & 0,2 \end{bmatrix}$$

$$\hat{\phi} = 0,8 \quad \hat{\Lambda}_x = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \quad \hat{\Theta}_\delta = \begin{bmatrix} 0,5 & \\ & 0,3 \end{bmatrix}$$

$$\hat{\psi} = 1$$

5.1.3 - Modelagem de Equações Simultâneas

Quando os construtos são formados apenas por variáveis observadas, ou manifestadas, os modelos são chamados de Análise de Caminhos ou Modelo de Equações Simultâneas. Usando a notação LISREL o modelo assume a seguinte estrutura:

$$Y_{p \times 1} = B_{p \times p} \cdot Y_{p \times 1} + \Gamma_{p \times q} \cdot X_{q \times 1} + \zeta_{p \times 1}$$

Onde $\eta = Y$ e $\xi = X$. As matrizes B e Γ contém os coeficientes dos efeitos diretos e a estrutura de covariância para o modelo é:

$$\Sigma(\theta) = \begin{bmatrix} (I - B)^{-1} (\Gamma \Phi \Gamma' + \Psi) (I - B)^{-1} & (I - B)^{-1} \Gamma \Phi \\ \Phi \Gamma' (I - B)^{-1} & \Phi \end{bmatrix}$$

$$\Sigma(\theta) = Cov \left(\begin{matrix} Y \\ X \end{matrix} \right) = \begin{bmatrix} Cov(Y) & Cov(YX) \\ Cov(XY) & Cov(X) \end{bmatrix}$$

Podemos modelar agora o exemplo da figura 5. Observe que nesse exemplo não existe relações entre variáveis endógenas para endógenas, portanto não existe a matriz B. Veja também que a matriz de covariância é igual à de correlação, pois os dados são padronizados.

Matriz de Covariância

$$\begin{matrix} & Y & X_1 & X_2 \\ Y & 1 & & \\ X_1 & 0,6 & 1 & \\ X_2 & 0,7 & 0,5 & 1 \end{matrix}$$

$$Cov(X) = \Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} = \begin{bmatrix} 1 & \\ 0,5 & 1 \end{bmatrix}$$

$$Cov(YX) = \Phi\Gamma' = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \times \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \end{bmatrix} = \begin{bmatrix} \phi_{11}\gamma_{11} + \phi_{12}\gamma_{12} \\ \phi_{21}\gamma_{11} + \phi_{22}\gamma_{12} \end{bmatrix} = \begin{bmatrix} 0,6 \\ 0,7 \end{bmatrix}$$

Onde obtemos o sistema de equações semelhante à equação (8)

$$\begin{cases} \gamma_{11} + 0,5\gamma_{12} = 0,6 \\ 0,5\gamma_{11} + \gamma_{12} = 0,7 \end{cases}$$

$$\gamma_{11} = 0,33$$

$$\gamma_{12} = 0,53$$

Podemos agora representar novamente o diagrama da figura 4

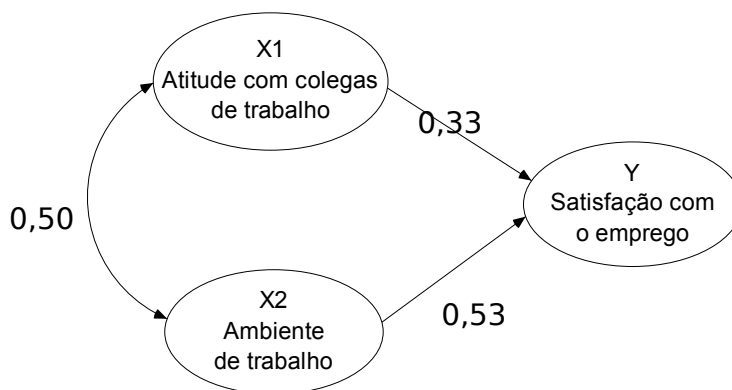


FIGURA 7 - Diagrama de caminhos obtido usando Modelagem de equações simultâneas

5.1.4 - Efeitos totais, diretos e indiretos

Os efeitos de uma variável sobre a outra pode ser medidos em três tipos: Totais, diretos e indiretos.

- Diretos: São os efeitos de uma variável diretamente sobre outra sem a interferência de nem uma outra variável no diagrama;
- Indiretos: É o efeito que está intermediado por pelo menos uma variável adicional;

- Totais: É a soma dos efeitos diretos e indiretos.

De Santos (2007) temos o seguinte resumo dos efeitos:

QUADRO 9 - Resumo dos efeitos diretos, indiretos e totais

Latentes (exógenas)				Latentes (endógenas)			
EFEITO	$\xi \rightarrow \eta$	$\xi \rightarrow Y$	$\xi \rightarrow X$	EFEITO	$\eta \rightarrow \eta$	$\eta \rightarrow Y$	$\eta \rightarrow X$
<i>Efeito direto</i>	Γ	0	Λ_x	<i>Efeito direto</i>	B	Λ_y	0
<i>Efeito indireto</i>	$(I - B)^{-1}\Gamma - \Gamma$	$\Lambda_y(I - B)^{-1}\Gamma$	0	<i>Efeito indireto</i>	$(I - B)^{-1} - I - B$	$\Lambda_y(I - B)^{-1} - \Lambda_y$	0
<i>Efeito total</i>	$(I - B)^{-1}\Gamma$	$\Lambda_y(I - B)^{-1}\Gamma$	Λ_x	<i>Efeito total</i>	$(I - B)^{-1} - I$	$\Lambda_y(I - B)^{-1}$	0

5.1.5 - Medidas de bondade de ajuste

Para a avaliação do ajuste dos modelos geralmente consideramos múltiplos índices de ajuste:

- A estatística de Qui-Quadrado (χ^2), seus graus de liberdade e seu nível de significância para H_0 : O modelo é ajustado;
- GFI (*goodness-of-fit*), índice que descreva a proporção de variância explicada do modelo, valores superiores a 0,9 indicam um bom ajuste do modelo;
- AGFI (*adjusted goodness of fit index*) índice que ajusta a proporção de variância explicada à complexidade do modelo, valores superiores a 0,9 indicam um bom ajuste do modelo;
- RMSA (*root mean square error of approximation*), índice de ajuste comparativo, valores inferiores a 0,08 são considerados aceitáveis;
- Qui-Quadrado normado (χ^2/df), índice de ajuste parcimonioso, valores inferiores a 3 são aceitáveis.

6 - Resultados

6.1 - Avaliação descritiva

6.1.1 - Indicadores de oferta

6.1.1.1 - Média de horas-aula diária

Para o município de João Pessoa a diferença de tempo em algumas escolas chega em 2005 a ser de uma hora e meia na 4° série e de duas horas para a 8° série . Apesar da obrigatoriedade de um de mínimo de 4 horas-aula por dia, ainda 15% das escolas ofereciam valores inferiores para a 4° série e perto de 35% para a 8° série, conforme visto nas Tabelas 5 e 6 em anexo.

6.1.1.2 - Porcentagem de docentes com curso superior

Apesar da não obrigatoriedade da formação superior dos professores nas escolas mais da metade delas tinham em 2005 mais de 70% dos seus professores com curso superior completo. Nas escolas com os anos finais apesar da obrigatoriedade de 100% para todas as escolas ainda 6% delas não possuem todos os professores com curso superior, porém, destes, a menor porcentagem é de 90%, ver Tabela 3 e 7 em anexo.

6.1.1.3 - Número de alunos por turma

Para o ano de 2005, João Pessoa possuía uma proporção média e mediana de aproximadamente 29 alunos por turma nos anos iniciais e de aproximadamente 34 alunos por turma para os anos finais, ver Tabela 3 e 4 em anexo.

6.1.2 - Indicadores de eficiência e rendimento

6.1.2.1 - Taxa de distorção idade-série

Em 2005 mais da metade das escolas tem mais que 50% de seus alunos fora da idade recomendada. Em algumas, a Taxa de Distorção chega a 80%, isto tanto para as escolas com anos iniciais quanto para as com anos finais. Observe Tabela 3 e 4.

6.1.2.2 - Taxa aprovação

Para os anos iniciais a *Taxa de Aprovação* média e mediana foi de 67%, para os anos finais a Taxa média e mediana foi próxima de 60%. Em ambos os casos a *Taxa de Aprovação* máxima estava na casa dos 80% não chegando aos 90%. Conforme Tabela 3 e 4.

6.1.3 - Notas do Prova Brasil ou Saeb

Para João Pessoa os anos iniciais têm em 2005 uma nota de Matemática com média e mediana próximos de 170 e com um desvio padrão de 11,14. Já a nota de Língua Portuguesa tinha valor médio e mediano próximo de 160 com desvio padrão de 11,99. Para os anos finais essas notas possuem média e mediana próximos de 220 para Matemática com desvio padrão de 9,46 e média e mediana próximos de 210 com desvio de 10,33 para Língua Portuguesa.

6.1.4 - IDEB

Em João Pessoa no ano de 2005 as escolas possuíam um valor médio e mediano do IDEB próximo de 2,7 para os anos iniciais com desvio padrão de 0,58 para os anos iniciais. Para os anos finais o valor médio e mediano era próximo de 2,4 com desvio padrão de 0,48.

6.2 - Modelagem de Equações Simultâneas

Como o SEM é uma técnica confirmatória e não exploratória, o modelo aborda a priori a seguinte temática: Os três indicadores de oferta causam os indicadores de eficiência e Rendimento e estas por sua vez causam as notas do Prova Brasil ou Saeb e o IDEB. Este modelo será aplicado aos anos iniciais e aos anos finais do ensino básico.

Dito isto construímos o seguinte sistema de equações:

QUADRO 10 - Relação das variáveis dependentes e independentes do modelo

Variáveis Dependentes	Variáveis independentes
Taxa de distorção idade-série	Média de horas-aula diária + Porcentagem de professores com formação superior + Número de alunos por turma
Taxa de aprovação	Média de horas-aula diária + Porcentagem de professores com formação superior + Número de alunos por turma
Matemática	Taxa de distorção idade-série + Taxa de aprovação
Língua Portuguesa	Taxa de distorção idade-série + Taxa de aprovação
IDEB	Taxa de distorção idade-série + Taxa de aprovação

Para este modelo montamos o seguinte diagrama de caminhos.

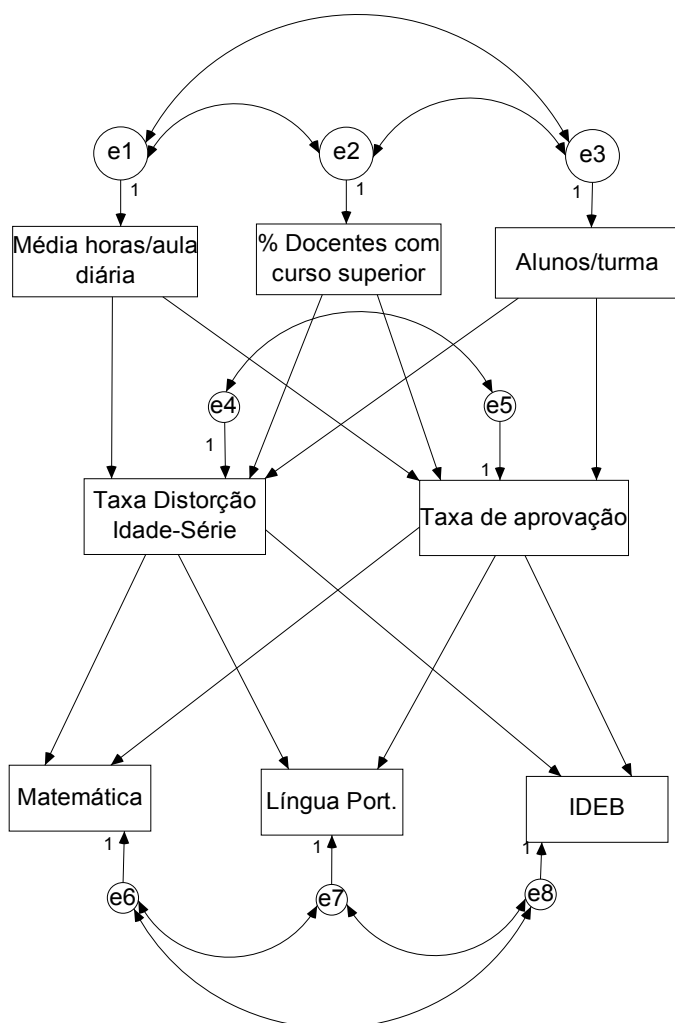


FIGURA 8 - Diagrama de caminhos do modelo sugerido

6.3 - Anos iniciais da educação básica

Dos resultados obtidos do Software observamos as significâncias das estimativas para a covariância entre as variáveis. Veja no Quadro 11 que as covariâncias não são significantes entre os Índices de oferta. Portanto estas covariâncias serão retiradas do modelo. Os Índices de oferta também não apresentam correlação entre si, veja Quadro 21 anexo. Porém as demais covariâncias são significantes no modelo.

QUADRO 11 - Estimativas para as covariâncias do modelo - anos iniciais

	Estimate	S.E.	C.R.	P
e1 <--> e2	-0,057	0,742	-0,076	0,939
e2 <--> e3	-2,474	13,254	-0,187	0,852
e1 <--> e3	0,046	0,165	0,280	0,780
e5 <--> e4	-55,642	15,498	-3,590	***
e7 <--> e6	73,265	14,032	5,221	***
e7 <--> e8	2,101	0,384	5,467	***
e6 <--> e8	2,108	0,375	5,619	***

*** P-valor menor que 0,001

Observando a significância dos coeficientes de regressão vemos que os coeficientes de *Porcentagem de professores com curso superior* para *Taxa de distorção idade-série* e para *Taxa de Aprovação* são não significantes, ou seja, não existe uma relação de causa e efeito entre elas. Vemos também esta não significância entre os indicadores de *Número de alunos/turma* para *Taxa de distorção idade-série* e *Taxa de Aprovação*. Sendo estas variáveis retiradas do modelo. Após retirarmos a relação entre *Taxa de aprovação* e *Língua Portuguesa*, não significativa a 5%, a relação entre a *Taxa de aprovação* e *Matemática* também se torna não significativa. Portanto estas relações também são retiradas do modelo

QUADRO 12 - Estimativas para os coeficientes de regressão do modelo - anos iniciais

			Estimate	S.E.	C.R.	P
y ₁	<---	x1	-29,551	5,745	-5,144	***
y ₂	<---	x1	10,427	3,668	2,843	0,004
y ₁	<---	x2	-0,072	0,072	-1,002	0,316
y ₂	<---	x2	-0,039	0,046	-0,857	0,392
y ₂	<---	x3	0,088	0,205	0,428	0,669
y ₁	<---	x3	0,173	0,322	0,537	0,591
y ₃	<---	y1	-0,24	0,082	-2,937	0,003
y ₅	<---	y1	-0,007	0,002	-3,216	0,001
y ₄	<---	y1	-0,303	0,085	-3,555	***
y ₃	<---	y2	0,287	0,142	2,026	0,043
y ₄	<---	y2	0,279	0,148	1,885	0,059
y ₅	<---	y2	0,048	0,004	12,759	***

*** P-valor menor que 0,001

O modelo final para os anos iniciais apresenta as seguintes características e medidas de bondade de ajuste:

QUADRO 13 - Características do modelo final e medidas de ajuste - anos iniciais

Tipo do modelo	Recursivo
Tamanho de amostra	73 (censo)
Método de estimação do modelo	Generalized Least Squares
Qui-Quadrado (χ^2)	10,576
Graus de liberdade (gl)	5
P-valor da estatística (χ^2)	0,06 valor desejado > 0,05
Qui-Quadrado normado (χ^2/gl)	2,115 valor desejado < 3
Índice de qualidade de ajuste (GFI)	0,951 valor desejado > 0,900
Índice ajustado de qualidade (AGFI)	0,794 valor desejado > 0,900
Raiz do erro quadrático médio de aproximação (RMSA)	0,120 valor desejado < 0,080

Apesar de algumas medidas estarem fora do desejável, o modelo foi considerado ajustado.

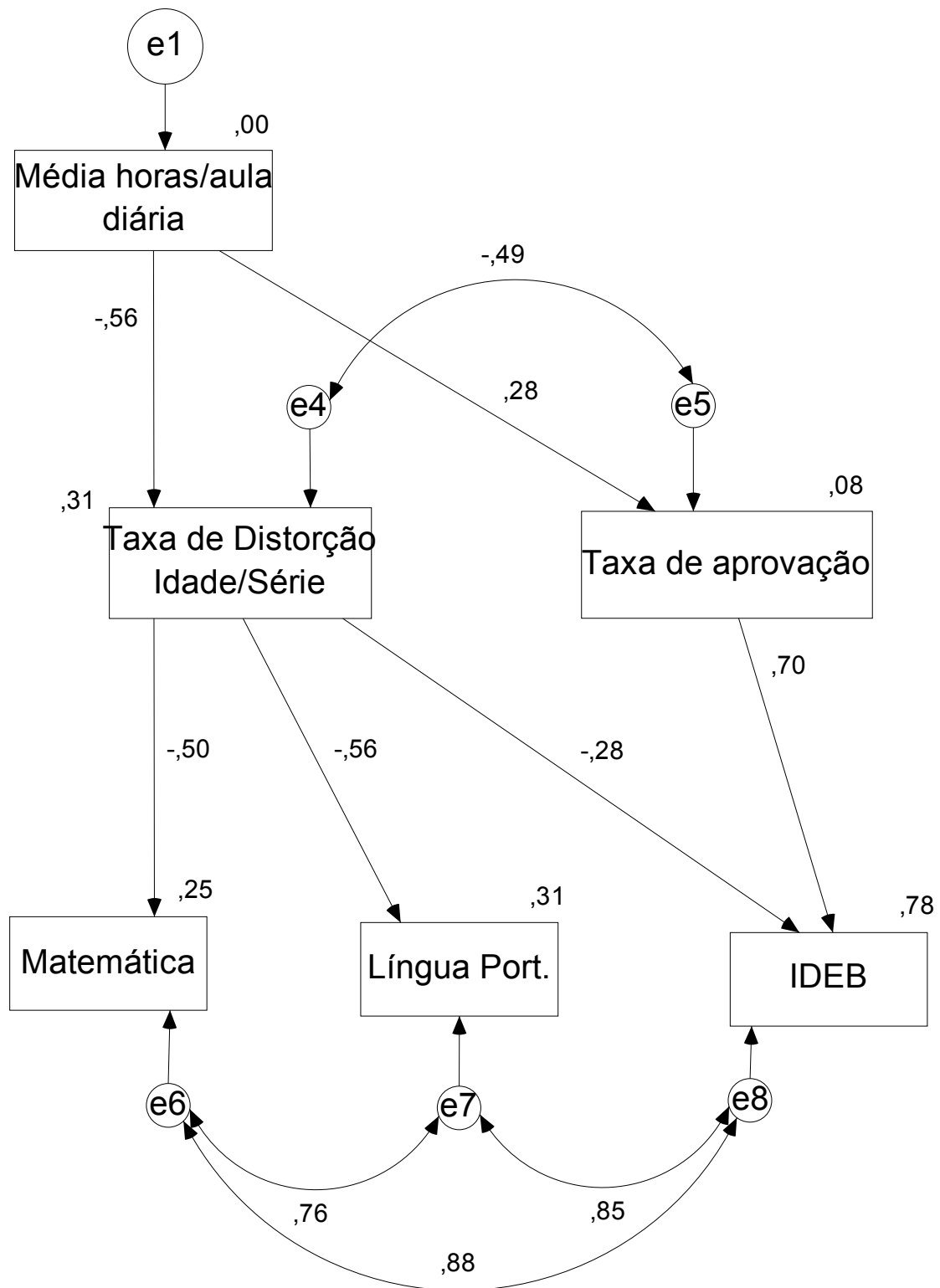


FIGURA 9 - Modelo final de equações simultâneas estimativas padronizadas - Anos iniciais

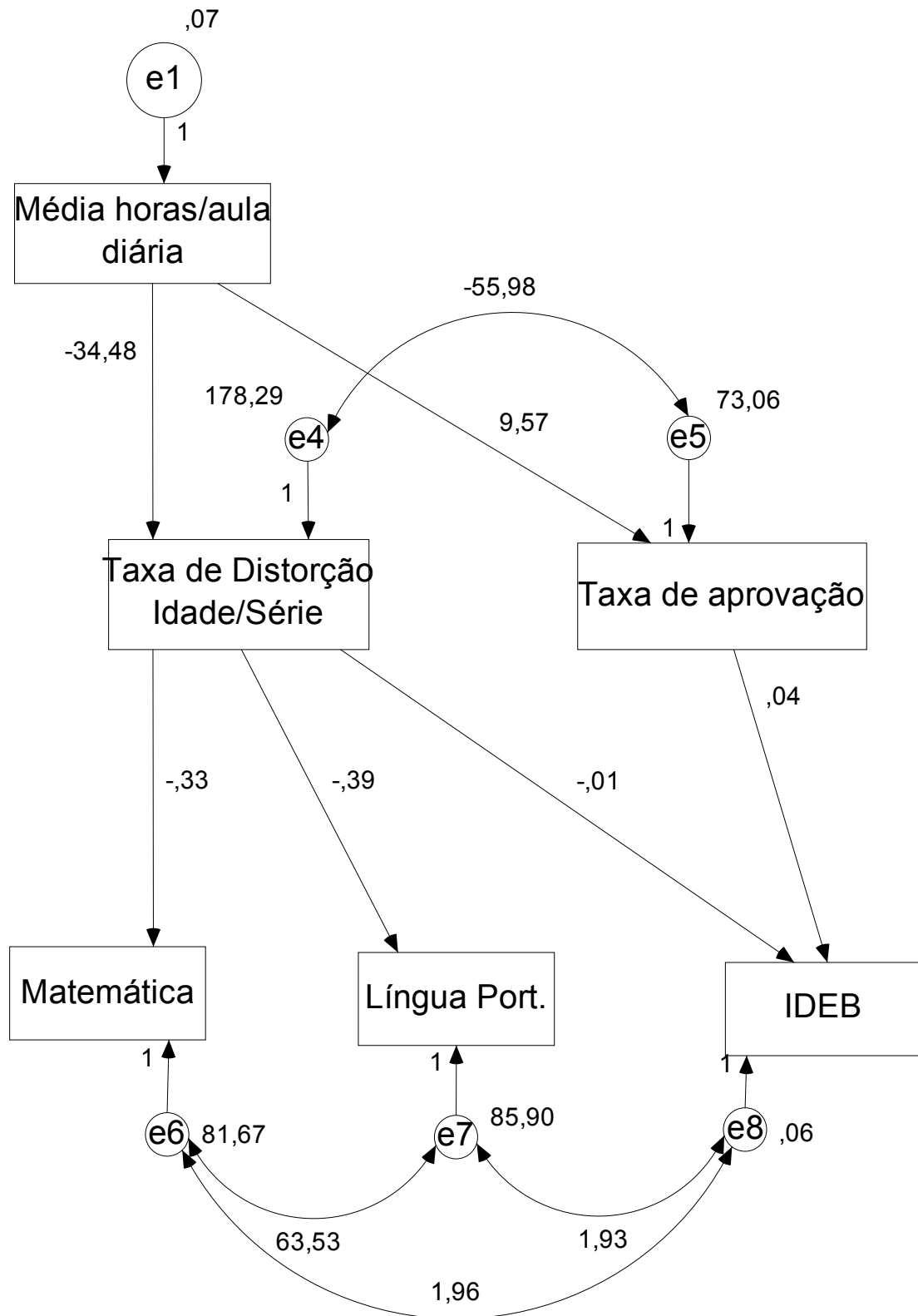


FIGURA 10 - Modelo final de equações simultâneas estimativas não padronizadas - Anos iniciais

A *Média de hora-aula* é responsável por um dos resultados mais importantes no modelo. Mostrando-se de grande importância para os índices de Eficiência e Rendimento. Observando o modelo padronizado vemos que a relação entre a *Média de hora-aula* é mais forte para a *Taxa de distorção* que para a *Taxa de Aprovação*. Pelo modelo não padronizado vemos que o aumento de uma hora-aula causa a redução de 34,48% na *Taxa de distorção* e o aumento de 9,57% na *Taxa de Aprovação*. Em contrapartida temos as variáveis *Porcentagem de professores com curso superior* e *Número de alunos/turma* que não se adequaram ao modelo.

A *Porcentagem de professores com curso superior* pode não ter se enquadrado devido ao fato que mais da metade das escolas já possuem mais que 50% dos seus professores com formação superior para os anos iniciais. Possuindo pouca variabilidade para participar do modelo. Porém ainda podemos observar-la em dois grupos, escolas com porcentagem menor que a mediana e escolas com porcentagem acima da mediana.

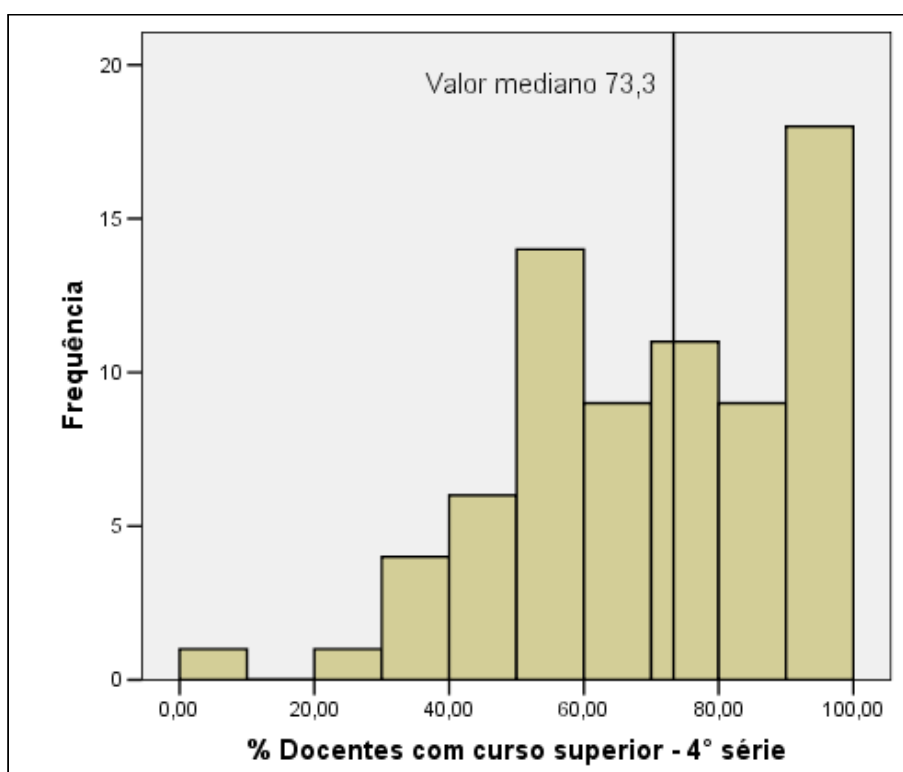


FIGURA 11 - Histograma de Porcentagens de docentes com curso superior

Observe que as escolas com valor abaixo da mediana se concentram em maior parte entre 40% e 70% caracterizando assim escolas com uma quantidade média de professores com curso superior. Vale lembrar a não obrigatoriedade legal da formação superior nos anos iniciais da educação básica, sendo assim podemos considerar as

escolas com porcentagem acima da mediana com alta quantidade de professores com curso superior. Comparando esses dois grupos entre as variáveis de Eficiência e Rendimento vemos:

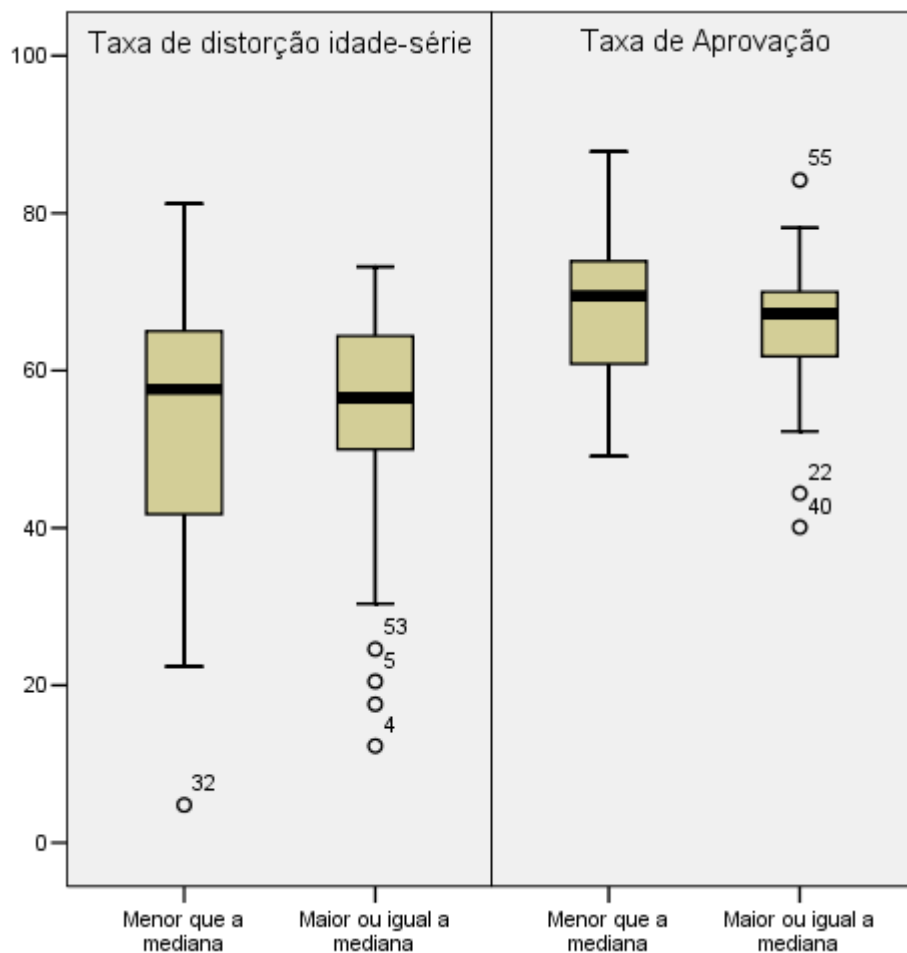


FIGURA 12 - Boxplot *Taxa de distorção idade-série* e *Taxa de Aprovação* por *Porcentagem de professores com curso superior*

Veja o teste de diferença de médias no quadro abaixo:

QUADRO 14 - Teste para amostras independentes

	Teste de Levene para variâncias iguais		Teste t para igualdade de médias		
	F	Sig.	t	gl	Sig. (bilateral)
% Distorção Idade/Série - 4° série	0,488	0,487	-0,002	71,000	0,998
			-0,002	70,715	0,998
<i>Taxa de Aprovação Séries Iniciais até 4ª série</i>	0,230	0,633	1,653	71,000	0,103
			1,653	70,842	0,103

Através da demonstração gráfica, apoiada pelo teste de comparação de médias vemos a igualdade dentre esses grupos, comprovamos que a diferença entre as escolas com uma quantidade média de professores com curso superior e as escolas com quantidade alta não foi significativa para as Taxas de distorção e Aprovação para o município de João Pessoa no ano de 2005.

O *Número de alunos por turma* nas escolas de João Pessoa em 2005 apresenta valores sempre bem próximos da média não existindo valores extremos que possam interferir nas demais variáveis. Como vemos no histograma abaixo.

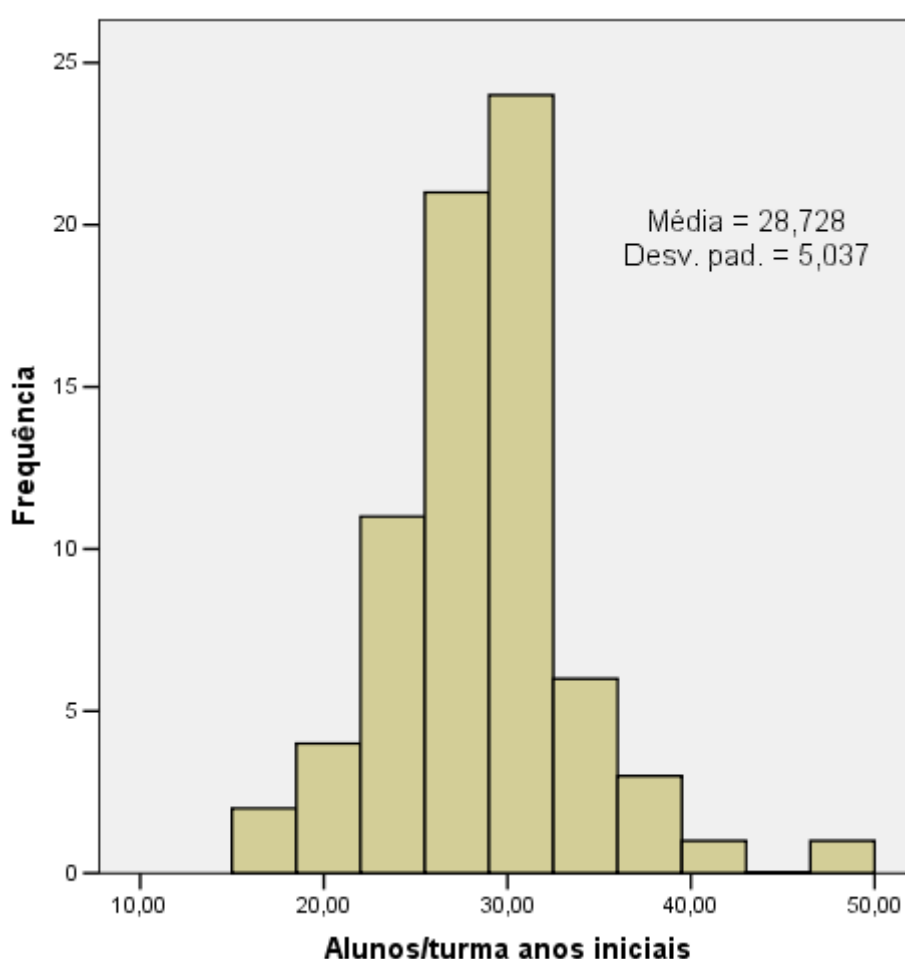


FIGURA 13 - Histograma alunos/turma anos iniciais

Vemos também as variáveis de Eficiência e Rendimento que, correlacionadas entre si, explicam as Variáveis de Qualidade, também correlacionadas, de forma que a *Taxa de distorção* causa um impacto negativo nas medidas de *Matemática* e *Língua Portuguesa* e um leve impacto no IDEB. Já a *Taxa de aprovação* possui uma relação

muito forte com o IDEB em João Pessoa no ano de 2005. Onde a variação de um desvio padrão na *Taxa de aprovação* causa uma variação de 0,7 desvio padrão no *IDEB*.

O quadro abaixo que contem os efeitos padronizados diretos, indiretos e totais entre as variáveis.

QUADRO 15 - Efeitos padronizados - anos iniciais

	Efeito	Média de horas-aula	Taxa de distorção	Taxa de aprovação
Taxa de distorção	Direto	-0,56		
	Indireto	0,00		
	Total	-0,56		
Taxa de aprovação	Direto	0,28		
	Indireto	0,00		
	Total	0,28		
Matemática	Direto	0,00	-0,50	
	Indireto	0,28	0,00	
	Total	0,28	-0,50	
Língua Portuguesa	Direto	0,00	-0,56	
	Indireto	0,31	0,00	
	Total	0,31	-0,56	
IDEB	Direto	0,00	-0,28	0,70
	Indireto	0,35	0,00	0,00
	Total	0,35	-0,28	0,70

Veja que a variável Média de horas aula possui efeito em todas as outras variáveis do modelo. Porém seu maior efeito indireto está no IDEB.

6.4 - Anos finais da educação básica

Dos resultados da saída do software vemos que o modelo se comporta exatamente da mesma maneira que o modelo para os anos iniciais, o que só confirma o modelo, já que as variáveis se comportam de maneira semelhante nos anos iniciais e nos anos finais do ensino básico.

Novamente as covariâncias entre os Índices de oferta são não significativas. Pelo Quadro 22 em anexo vemos que estas também não se correlacionam. Por isso são removidos do modelo para os anos finais.

QUADRO 16 - Estimativas para as covariâncias do modelo - anos finais

	Estimate	S.E.	C.R.	P
e1 <--> e2	-,025	,082	-,304	,761
e2 <--> e3	,030	1,295	,023	,982
e1 <--> e3	,118	,294	,402	,688
e5 <--> e4	-37,560	17,933	-2,094	,036
e6 <--> e7	41,891	12,229	3,425	***
e7 <--> e8	1,075	,266	4,038	***
e6 <--> e8	,949	,237	3,998	***

Observando a significância dos coeficientes de regressão vemos que as mesmas relações retiradas no modelo para os anos iniciais deverão ser retiradas no modelo para os anos finais. *Porcentagem de professores com curso superior* e *Número de alunos por turma* não conseguem explicar a *Taxa de distorção* nem a *Taxa de aprovação*. Bem como as relações entre a *Taxa de aprovação* e a nota de *Matemática* e a nota de *Língua Portuguesa* não são significativas e, como as demais relações, são retiradas do modelo.

QUADRO 17 - Estimativas para os coeficientes de regressão do modelo - anos finais

	Estimate	S.E.	C.R.	P
y ₁ <--- x ₁	-23,941	5,378	-4,452	***
y ₂ <--- x ₁	9,091	3,989	2,279	,023
y ₁ <--- x ₂	-,372	1,329	-,280	,779
y ₂ <--- x ₂	1,422	,980	1,452	,146
y ₁ <--- x ₃	-,570	,352	-1,619	,105
y ₂ <--- x ₃	,188	,261	,721	,471
y ₃ <--- y ₁	-,272	,089	-3,065	,002
y ₄ <--- y ₁	-,282	,099	-2,850	,004
y ₅ <--- y ₁	-,005	,002	-3,014	,003
y ₃ <--- y ₂	,027	,136	,195	,845
y ₄ <--- y ₂	-,013	,152	-,085	,932
y <--- y	,040	,003	14,607	***

		Estimate	S.E.	C.R.	P
5	2				

O modelo final para os anos finais apresenta as seguintes características e medidas de bondade de ajuste:

QUADRO 18 - Características do modelo final e medidas de ajuste - anos finais

Tipo de modelo	Recursivo
Tamanho de amostra	49 (censo)
Método de estimação do modelo:	Generalized Least Squares
Qui-Quadrado (χ^2)	3,138
Graus de liberdade (<i>g.l.</i>)	5
	0,67
P-valor da estatística (χ^2)	9 valor desejado > 0,05
	0,62
Qui-Quadrado normado ($\chi^2/g.l.$)	8 valor desejado < 3
	0,97
Índice de qualidade de ajuste (GFI)	8 valor desejado > 0,900
	0,90
Índice ajustado de qualidade (AGFI)	8 valor desejado > 0,900
	0,00
Raiz do erro quadrático médio da aproximação (RMSA)	0 valor desejado < 0,080

O modelo foi considerado ajustado, pois apresenta o valor desejado para todas as medidas de ajuste. Estes mostram que as relações presentes são melhores observadas para os anos finais.

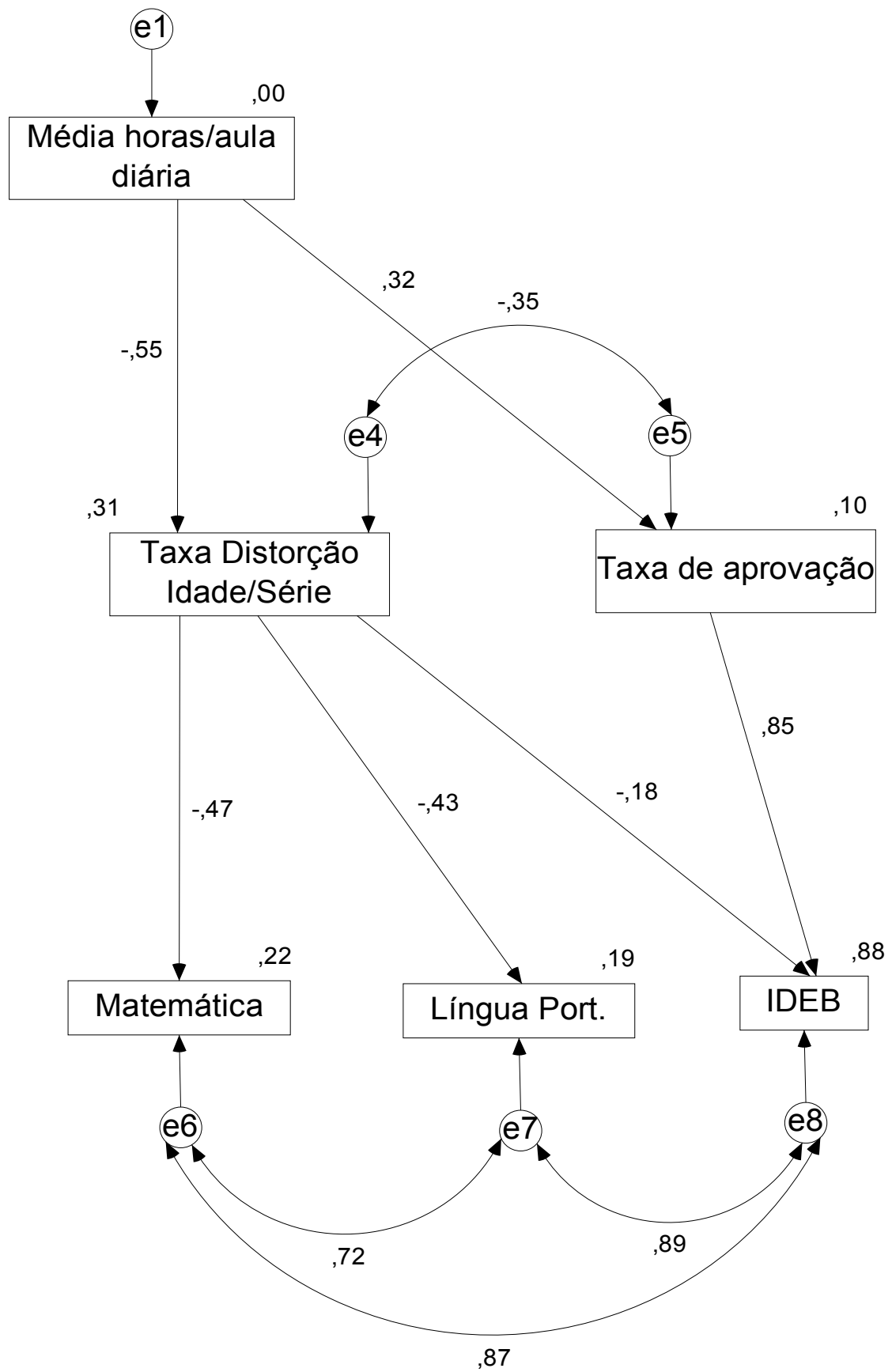


FIGURA 14 - Modelo final de equações simultâneas estimativas padronizadas - Anos finais

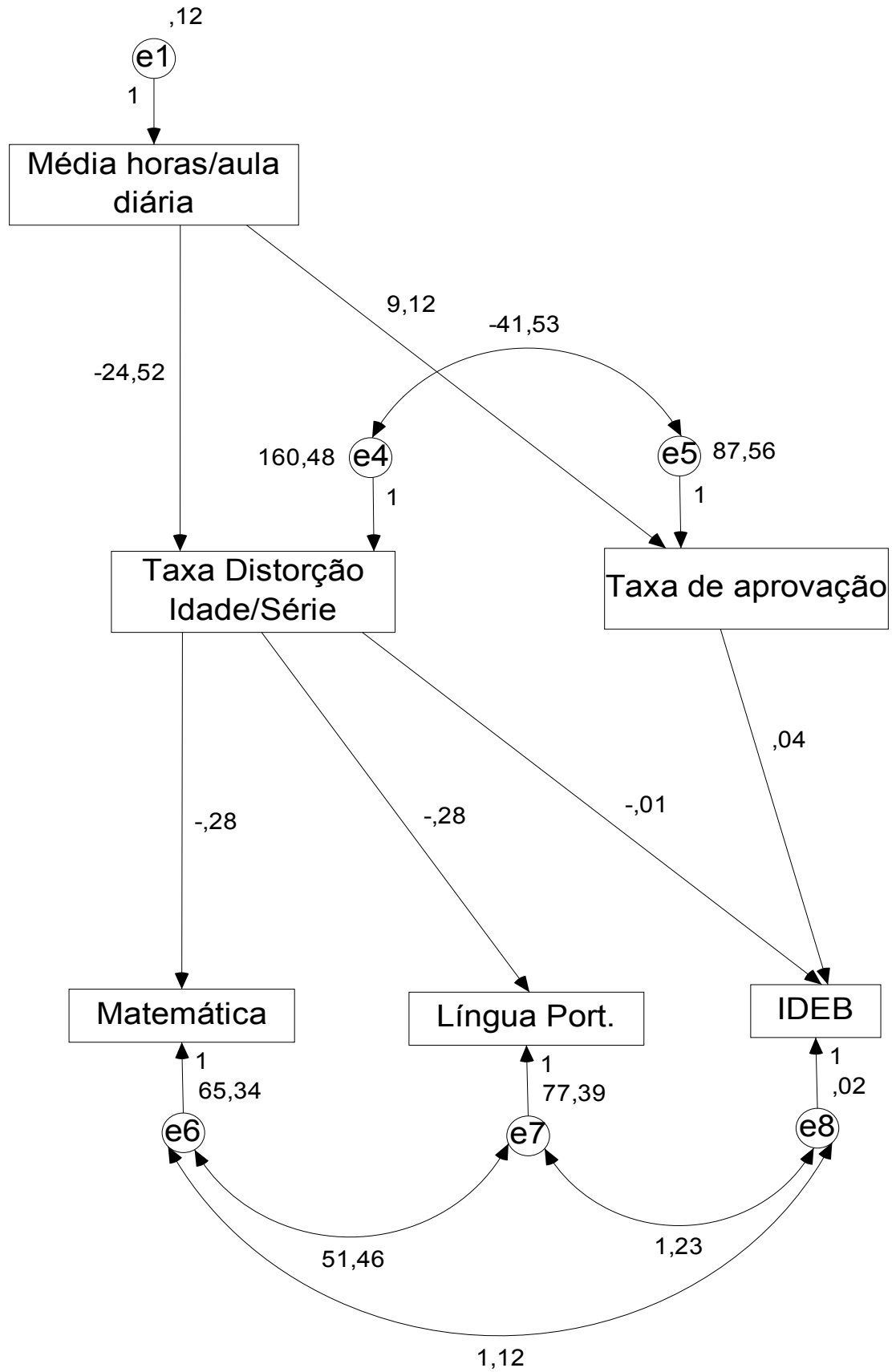


FIGURA 15 - Modelo final de equações simultâneas estimativas não padronizadas - Anos finais

Para os anos finais a *Média de hora-aula* causa influencia nas Taxas de Distorção e de Aprovação. Sendo que, destas relações, a com *Taxa de distorção* é a mais forte. Em João Pessoa para o ano de 2005 o aumento de uma hora aula causa a diminuição de 24,52% na *Taxa de distorção* e o aumento de 9,12% na *Taxa de aprovação*.

A *Porcentagem de professores com curso superior* não se enquadra no modelo devido a sua falta de variabilidade. O que é compreensível devido a obrigatoriedade de das escolas de possuírem 100% dos seus professores para os anos finais com curso superior.

Veja no Quadro abaixo que o coeficiente de variação dos anos finais da educação básica é maior que o dos anos iniciais. Apesar da variação no *Número de alunos por turma* ser maior para os anos finais esta variação ainda não é suficiente para estudar um impacto nas demais variáveis do modelo.

QUADRO 19 - Coeficiente de variação dos conjuntos de dados

	Coeficiente de variação
Anos iniciais	17,54%
Anos finais	17,73%

Nas relações entre as variáveis de Eficiência e Rendimento e as variáveis de Qualidade observamos que a *Taxa de distorção* causa um efeito negativo nas variáveis *Matemática* e *Língua Portuguesa* e um pequeno efeito no IDEB já a *Taxa de aprovação* não consegue se relacionar com as variáveis *Matemática* e *Língua Portuguesa*, porém é bastante explicativa para o IDEB.

O quadro abaixo que contem os efeitos padronizados diretos, indiretos e totais entre as variáveis.

QUADRO 20 - Efeitos padronizados - anos finais

	Efeito	Média de horas-aula	Taxa de distorção	Taxa de aprovação
Taxa de distorção	Direto	-0,55		
	Indireto	0,00		
	Total	-0,55		
Taxa de aprovação	Direto	0,32		
	Indireto	0,00		
	Total	0,32		
Matemática	Direto	0,00	-0,47	
	Indireto	0,26	0,00	
	Total	0,26	-0,47	
Língua Portuguesa	Direto	0,00	-0,43	
	Indireto	0,24	0,00	
	Total	0,24	-0,43	
IDEB	Direto	0,00	-0,18	0,85
	Indireto	0,37	0,00	0,00
	Total	0,37	-0,18	0,85

A Média de horas-aula além de possuir efeito direto nos Índices de eficiência possui um efeito indireto nas notas de Matemática e Língua Portuguesa, principalmente no IDEB onde seu efeito indireto é mais forte.

7 - Conclusões

A utilização da Modelagem de Equações Simultâneas foi capaz de perceber as relações entre as variáveis apresentadas no estudo. Respondendo ao objetivo de descobrir que variáveis explicam a qualidade, aqui representada pelas medidas usadas pelo próprio governo, dentre as variáveis estudadas. Descobrimos que a Taxa de Aprovação é um grande indicador de qualidade para o município de João Pessoa, devido a sua forte ligação com o IDEB. Os resultados indicam que onde as crianças aprendem e passam de ano o Índice de Desenvolvimento da Educação Básica também será alto para todos os anos da educação básica.

Descobrimos também a ligação entre a Taxa de distorção idade-série e os resultados da prova de Matemática e Língua Portuguesa realizados pelo Prova Brasil ou Saeb. A Taxa de distorção age de forma negativa nessas duas avaliações. Os alunos que ficam reprovados, além de não ter aprendido, acabam ficando desmotivados e são responsáveis por piores resultados nas provas avaliativas.

Sabendo do impacto que essas duas medidas de Eficiência e Rendimento causam nos índices de qualidade, descobrimos que a média de horas-aula diária é diretamente influente nessas medidas e, portanto, indiretamente influente nas medidas de qualidade. O aumento do tempo que os alunos passam nas escolas aumenta o rendimento escolar e conseqüentemente os resultados nas avaliação de qualidade.

Não é desejável um sistema de ensino que reprove constantemente os alunos levando-os muitas vezes a evasão escolar, porém um sistema que aprove todos os alunos independente da aprendizagem apenas maquiará resultados para as estatísticas educacionais. Os resultados deste trabalho indicam duas soluções para a melhoria da qualidade do ensino: O aumento do número de horas-aula nas escolas ,conseqüentemente diminuindo a Taxa de distorção e aumentando a Taxa de aprovação, possivelmente até a adoção do ensino em tempo integral e utilização de medidas corretivas da distorção dos alunos como turmas de aceleração.

8 - Referencias Bibliograficas

BRASIL. Constituição Federal da República Federativa do Brasil. **Diário Oficial da União**. Brasília, 5 de outubro de 1988.

_____. Lei nº 9.394, de 20 de dezembro de 1996. Lei de Diretrizes e Bases da Educação (LDB). **Diário Oficial da União**. Brasília, 23 de dezembro de 1996.

_____. Lei nº 9.424, de 24 de dezembro de 1996. Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério (FUNDEF). **Diário Oficial da União**. Brasília, 26 de dezembro de 1996.

_____. Lei nº 10.172, de 9 de janeiro de 2001. Plano Nacional de Educação (PNE). **Diário Oficial da União**. Brasília, Ministério da Educação, 2001.

HAIR, J. F. et al. **Análise multivariada de dados.**, 5ª ed. - São Paulo: Bookman Editora, 2005.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. Englewood Cliffs, New Jersey: Prentice-Hall, 1982.

MEC/INEP. **Levantamento do custo-aluno-ano em escolas de Educação Básica que oferecem condições de oferta para um ensino de qualidade – 2ª Etapa**. Relatório Nacional da Pesquisa. Salvador, 2005.

_____. **Pesquisa nacional de qualidade na educação**. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2006. 2 v.

_____. **Indicadores da qualidade na educação**. 3ª ed. ampliada. Ação Educativa, Unicef, Pnud, INEP, Seb/MEC (coord.) – São Paulo: Ação Educativa, 2007. 72 p.

_____. **Dicionário de Indicadores Educacionais**. Fórmulas de Cálculo. Coordenação-Geral de Sistemas Integrados de Informações Educacionais. Brasília, fevereiro de 2004.

MENEZES, Ebenezer Takuno de; SANTOS, Thais Helena dos. "Legislação educacional" (verbete). **Dicionário Interativo da Educação Brasileira** - EducaBrasil. São Paulo: Midiamix Editora, 2002. Disponível em: <<http://www.educabrasil.com.br/eb/dic/dicionario.asp?id=174>>. Acesso em 23 de abril de 2008.

PARO, V. H. A gestão da educação ante as exigências de qualidade e produtividade das escolas pública. In: SILVA, L. H. da (Org). **A escola cidadã no contexto da globalização**. Petrópolis: Vozes, 1998.

SANTOS, JOZEMAR PEREIRA DOS. **Seminário do depto. Estatística: Modelagem de Equações Estruturais**. João Pessoa: Depto. Estatística CCEN/UFPB, 2007. 60 slides, color.

ANEXO

TABELA 3 - Medidas descritivas das variáveis da educação básica - anos iniciais

	Média horas/aula diária - 4º série	Porcentagem de docentes com curso superior - 4º série	Alunos/turma anos iniciais	Taxa de distorção Idade/Série - 4º série	Taxa de aprovação - Séries Iniciais até 4ª série	Matemática 4ª série	Língua Portuguesa a 4ª série	IDEB - Anos iniciais
Mínimo	3,00	0,00	17,56	4,80	40,10	149,40	136,50	1,50
Média	4,12	70,18	28,73	52,67	67,12	171,18	162,62	2,81
Mediana	4,10	73,30	28,75	56,50	67,60	170,00	161,70	2,70
Máximo	4,50	100,00	46,55	81,20	87,80	212,70	198,90	4,70
Amplitude Total	1,50	100,00	28,99	76,40	47,70	63,30	62,40	3,20
Desvio Padrão	0,28	22,63	5,04	16,17	9,32	11,14	11,99	0,58

Fonte: Saeb 2005, Censo Escolar 2005 e 2006, INEP, Secretaria de educação do município de João Pessoa

TABELA 4 - Medidas descritivas das variáveis da educação básica - anos finais - ano 2005

	Média horas/aula diária - 8º série	Porcentagem de docentes com curso superior - 8º série	Alunos/turma anos finais	Taxa de distorção Idade/Série - 8º série	Taxa de aprovação 5ª a 8ª série	Matemática 8ª série EF	Língua Port. 8ª série EF	IDEB - Anos finais
Mínimo	3,40	90,00	17,67	36,40	42,00	201,50	176,20	1,50
Média	4,10	99,61	33,61	64,90	60,40	223,74	210,33	2,38
Mediana	4,00	100,00	33,88	65,10	59,60	224,20	212,70	2,40
Máximo	5,40	100,00	43,93	91,20	86,80	242,50	234,30	3,70
Amplitude Total	2,00	10,00	26,26	54,80	44,80	41,00	58,10	2,20
Desvio Padrão	0,36	1,69	5,96	15,37	10,00	9,46	10,33	0,48

Fonte: Saeb 2005, Censo Escolar 2005 e 2006, INEP, Secretaria de educação do município de João Pessoa

TABELA 5 - Distribuição de frequência média horas/aula diária - 4° série

Média horas-aula diária	Frequência	Percentual acumulado
3,00	1	1,37
3,50	2	4,11
3,60	2	6,85
3,70	2	9,59
3,80	3	13,70
3,90	1	15,07
4,00	16	36,99
4,10	12	53,42
4,20	12	69,86
4,30	8	80,82
4,40	2	83,56
4,50	12	100,00
Total	73	

TABELA 6 - Distribuição de frequência média horas/aula diária - 8° série

Média horas-aula diária	Frequência	Percentual acumulado
3,40	1	2,04
3,70	2	6,12
3,80	7	20,41
3,90	7	34,69
4,00	10	55,10
4,10	5	65,31
4,20	8	81,63
4,30	3	87,76
4,40	1	89,80
4,50	1	91,84
4,70	1	93,88
5,00	1	95,92
5,20	1	97,96
5,40	1	100,00
Total	49	

TABELA 7 - Distribuição de frequência de porcentagem de professores com curso superior - anos finais

	Frequência	Percentual	Percentual Cumulativo
90,00	1	2,04	2,04
94,40	1	2,04	4,08
96,40	1	2,04	6,12
100,00	46	93,88	100,00
Total	49	100,00	

QUADRO 21 - Matriz de correlação das variáveis - anos iniciais

	Média horas-aula diária - 4° série	Porcentagem de Docentes com curso superior - 4° série	Alunos/turma anos iniciais	Taxa de Distorção Idade/Série - 4° série	Taxa de aprovação Séries Iniciais até 4ª série	Matemática 4ª série EF	Língua Port. 4ª série EF	IDEB - anos finais
Média horas-aula diária - 4° série	1,00	-0,01	0,03	-0,51	0,32	0,37	0,44	0,38
		0,94	0,78	0,00	0,01	0,00	0,00	0,00
Porcentagem de Docentes com curso superior - 4° série	-0,01	1,00	-0,02	-0,10	-0,10	0,00	0,04	-0,09
	0,94		0,86	0,42	0,40	0,99	0,76	0,44
Alunos/turma anos iniciais	0,03	-0,02	1,00	0,04	0,06	0,17	0,10	0,11
	0,78	0,86		0,74	0,61	0,16	0,38	0,35
Taxa de Distorção Idade/Série - 4° série	-0,51	-0,10	0,04	1,00	-0,53	-0,47	-0,52	-0,61
	0,00	0,42	0,74		0,00	0,00	0,00	0,00
Taxa de aprovação Séries Iniciais até 4ª série	0,32	-0,10	0,06	-0,53	1,00	0,42	0,43	0,88
	0,01	0,40	0,61	0,00		0,00	0,00	0,00
Matemática 4ª série	0,37	0,00	0,17	-0,47	0,42	1,00	0,84	0,76
	0,00	0,99	0,16	0,00	0,00		0,00	0,00
Língua Port. 4ª série	0,44	0,04	0,10	-0,52	0,43	0,84	1,00	0,75
	0,00	0,76	0,38	0,00	0,00	0,00		0,00
IDEB - anos iniciais	0,38	-0,09	0,11	-0,61	0,88	0,76	0,75	1,00
	0,00	0,44	0,35	0,00	0,00	0,00	0,00	

Células contêm: Correlação de Pearson e Significância bilateral

Correlação Significante a 0,01 (bilateral)

QUADRO 22 - Matriz de correlação das variáveis - anos finais

	Média horas-aula diária - 8° série	Porcentagem de Docentes com curso superior - 8° série	Alunos/turma anos finais	Taxa de Distorção Idade/Série - 8° série	Taxa de aprovação 5° a 8ª série	Matemática 8ª série	Língua Port. 8ª série	IDEB - anos finais
Média horas-aula diária - 8° série	1,00	-0,07	0,14	0,30	-0,53	0,38	0,40	0,42
		0,61	0,35	0,04	0,00	0,01	0,00	0,00
Porcentagem de Docentes com curso superior - 8° série	-0,07	1,00	0,05	0,18	-0,01	0,03	0,04	0,13
	0,61		0,75	0,22	0,95	0,84	0,76	0,38
Alunos/turma anos finais	0,14	0,05	1,00	0,11	-0,22	0,35	0,32	0,23
	0,35	0,75		0,43	0,13	0,01	0,03	0,12
Taxa de Distorção Idade/Série - 8° série	0,30	0,18	0,11	1,00	-0,45	0,23	0,18	0,92
	0,04	0,22	0,43		0,00	0,12	0,22	0,00
Taxa de aprovação 5° a 8ª série	-0,53	-0,01	-0,22	-0,45	1,00	-0,45	-0,41	-0,55
	0,00	0,95	0,13	0,00		0,00	0,00	0,00
Matemática 8ª série	0,38	0,03	0,35	0,23	-0,45	1,00	0,79	0,55
	0,01	0,84	0,01	0,12	0,00		0,00	0,00
Língua Port. 8ª série	0,40	0,04	0,32	0,18	-0,41	0,79	1,00	0,51
	0,00	0,76	0,03	0,22	0,00	0,00		0,00
IDEB - anos finais	0,42	0,13	0,23	0,92	-0,55	0,55	0,51	1,00
	0,00	0,38	0,12	0,00	0,00	0,00	0,00	

Células contêm: Correlação de Pearson e Significância bilateral

Correlação Significante a 0,05 (bilateral)