

## CONSTRUTOS MENSURADOS NO CENSO ESCOLAR 2002 - ENSINO FUNDAMENTAL

Marcos Ruben de Oliveira y Jacob Arie Laros

### 1. INTRODUÇÃO

O Censo Escolar coleta dados de todas as escolas que compõem a rede de educação básica brasileira. É realizado anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep, autarquia federal vinculada ao Ministério da Educação, em parceria com as Secretarias de Educação dos Estados e do Distrito Federal. Segundo informações do próprio Inep,

*“O Censo Escolar é um levantamento de informações estatístico-educacionais de âmbito nacional, realizado anualmente. Esse levantamento abrange a educação básica, em suas diferentes etapas/níveis – educação infantil, ensino fundamental e Ensino Fundamental – e modalidades – ensino regular, educação especial e educação de jovens e adultos e também educação profissional de nível técnico” (Inep, 2006).*

A riqueza de dados oferecida pelo Censo Escolar permite inúmeras análises, como por exemplo: o Mapa do Analfabetismo no Brasil (Inep, 2003), a Geografia da Educação Brasileira (Inep, 2000), e O censo escolar (Diniz, 1999). Composto de uma base de dados de mais de 1.000 variáveis, o Censo Escolar 2002 está dividido em 11 blocos: 1) cadastro; 2) caracterização física; 3) dados gerais; 4) educação infantil; 5) classe de alfabetização; 6) ensino fundamental; 7) Ensino Fundamental; 8) curso normal em nível médio; 9) educação especial; 10) educação de jovens e adultos e 11) educação profissional. No entanto, essa mesma riqueza gera um complicador: a grande quantidade de variáveis no banco de dados dificulta a maioria das análises que podem ser feitas com base nesses dados. Algumas informações coletadas podem estar altamente associadas entre si, ou mesmo serem redundantes, e resultados com excesso de variáveis dificultam tanto a interpretação quanto o uso prático.

Em situações em que a quantidade de variáveis pode dificultar as análises, é extremamente útil fazer uso de técnicas estatísticas capazes de agrupar variáveis que tenham associação entre si, compondo pequenos grupos de componentes. O objetivo do presente estudo é identificar um pequeno número de componentes que represente as várias características das escolas de Ensino Fundamental avaliadas no Censo Escolar 2002. Segundo Laros (2005), “O objetivo da redução de dados é descobrir ponderações ótimas para 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”, e a Análise de Componentes Principais (ACP) será a técnica estatística utilizada para alcançar este fim.

### 2. MÉTODO

#### 2.1. Fonte de dados

O banco de dados utilizados nesse estudo é o do Censo Escolar 2002. A coleta de dados, planejada e executada pelo Inep em colaboração com as secretarias estaduais de educação, utilizou um questionário dividido em 11 blocos, compondo um total de mais de 1.000 questões. Esse instrumento

cumpre várias funções na gestão pública da educação, como por exemplo, fornecer informações para determinar a distribuição das verbas do FUNDEF - Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério, implantado nacionalmente em janeiro de 1998, cuja maior inovação consistiu na mudança da estrutura de financiamento do ensino fundamental no País, ao subvincular a esse nível de ensino uma parcela dos recursos constitucionalmente destinados à educação.

TABELA 1. DISTRIBUIÇÃO DAS ESCOLAS E ALUNOS DO ENSINO FUNDAMENTAL POR DEPENDÊNCIA ADMINISTRATIVA E LOCALIZAÇÃO EM 2002

Dependência Administrativa	Escolas					
	Rural		Urbana		Total	
	N	%	N	%	N	%
Estadual	7.627	8%	24.689	35%	32.316	19%
Federal	7	0%	37	0%	44	0%
Municipal	93.300	92%	28.036	39%	121.336	70%
Particular	402	0%	18.410	26%	18.812	11%
Total (escolas)	101.336	100%	71.172	100%	172.508	100%

Dependência Administrativa	Alunos					
	Rural		Urbana		Total	
	N	%	N	%	N	%
Estadual	724.319	11%	13.511.701	47%	14.236.020	41%
Federal	545	0%	25.877	0%	26.422	0%
Municipal	5.555.089	88%	12.098.054	42%	17.653.143	50%
Particular	38.825	1%	3.195.952	11%	3.234.777	9%
Total (alunos)	6.318.778	100%	28.831.584	100%	35.150.362	100%

Em 2002, 256.986 escolas ofereciam a educação básica, das quais 172.508 ofereciam o Ensino Fundamental e formam o escopo deste estudo. A distribuição dessas escolas por rede de ensino e tipo de localização revela o perfil da política educacional brasileira: as escolas que oferecem o Ensino Fundamental são, em sua grande maioria, escolas municipais, representando 70% do total de escolas deste nível de ensino (ver Tabela 1). O mesmo tipo de distribuição pode ser observado quando se trata do número de alunos matriculados: concentração na rede municipal, particularmente na zona urbana.

## 2.2. Variáveis consideradas na análise

A inspeção das perguntas dos 11 blocos do questionário do Censo Escolar 2002 revela que alguns grupos de questões apresentam grandes similaridades entre si, o que sugere que seja feita uma seleção criteriosa das variáveis do Censo antes de se aplicar a ACP propriamente dita, com o objetivo de evitar redundâncias óbvias. Como o foco deste estudo é o ensino fundamental, utilizamos apenas os blocos 1 (cadastro), 2 (caracterização física), 3 (dados gerais) e 6 (ensino fundamental). Do bloco 1, utilizou-se somente as variáveis de identificação das escolas; do bloco 2, as questões referentes aos equipamentos, dependências e recursos disponíveis na escola; e do bloco 3, a formação do professor que leciona para o ensino fundamental. Esta definição de escopo de análise reduz significativamente a quantidade de variáveis envolvidas, de tal forma que das aproximadamente 1.000 variáveis originais passamos para cerca de 400.

Outra redução no número de variáveis envolvidas na análise provém do fato de que, no Censo Escolar, algumas informações são abordadas mais de uma vez sobre diferentes enfoques. Assim, por exemplo, sobre o tema ‘alunos matriculados no Ensino Fundamental’ pergunta-se pela quantidade de alunos por turma, por gênero, por ano de nascimento, por série que está frequentando e pelo seu movimento na rede escolar (aprovado, reprovado, transferido, etc.). Estas formas de questionamento dão origem a vários indicadores escolares (Inep, 2004) que resumem algumas dezenas de variáveis. Assim, no presente estudo, optamos por usar os seguintes indicadores: percentual de alunos matriculados no turno noturno; número médio de alunos por turma; percentual de alunos do sexo feminino; taxa de distorção idade-série; taxa de aprovação no ano anterior; média de hora-aula diária e idade mediana dos alunos no ano. Por fim, agrupamos variáveis muito semelhantes com o objetivo de evitar redundâncias óbvias:

- O Censo Escolar questiona sobre a quantidade de computadores disponíveis na escola, discriminando três tipos de computadores. Na presente análise optou-se por tratar a quantidade de computadores sem diferenciação entre os tipos, utilizando o indicador “quantidade de computadores por aluno”.
- Oito questões do Censo Escolar são relacionadas com a utilização dos equipamentos de informática nas escolas, sendo que apenas a questão “A escola está ligada à Internet” será utilizada, pois se observou que esta última apresenta maior correlação com todas as demais, o que a torna melhor representante do conjunto de variáveis.
- Em relação ao uso de energia elétrica, água e esgoto, embora seja questionado vários aspectos a respeito, utilizou-se apenas os itens referentes ao uso do recurso público para esses fins.
- Das cinco perguntas sobre o tratamento do lixo, a que apresentou maior variabilidade foi a pergunta “Faz coleta periódica de lixo”, sendo também a que apresentou maior correlação com as demais questões sobre o mesmo tema e foi a variável selecionada para permanecer na análise.

O conjunto de decisões descrito acima resultou num banco de dados composto por 54 variáveis de interesse, que foram submetidas às análises seguintes.

### 3. RESULTADOS

#### 3.1. Análise exploratória e verificação da normalidade

Foram feitas algumas análises exploratórias para descobrir possíveis problemas no banco de dados, além de se calcular estatísticas exploratórias, com o objetivo de detectar *outliers* univariados. Neste sentido, observou-se a medida de assimetria, seguida de análise detalhada dos casos em que esta estatística assumiu valores maiores que 10. Assim, das 172.508 escolas iniciais, detectou-se 48 escolas do ensino fundamental com algum *outlier* univariado, que foram excluídas da análise.

A ACP não assume suposições quanto à distribuição das variáveis, no entanto a existência de normalidade nos dados faz com que a estrutura da solução seja realçada (Tabachnick y Fidell, 2001). Osborne (2002) analisa o efeito de algumas transformações, como a aplicação do logaritmo, o cálculo da raiz quadrada e o cálculo da inversa, com o objetivo de tornar as distribuições das variáveis mais aderentes à suposição de normalidade. Ao testar essas transformações nas variáveis quantidade de retroprojetores, copiadoras, vídeo-cassetes, aparelhos de fax, antenas parabólicas e o indicador média

de computadores por alumnos, constatou-se que a transformação que mais reduziu a assimetria foi o cálculo da raiz quadrada, ao passo que para as variáveis quantidade de TVs, impressoras, aparelhos de som, ventiladores, alunos matriculados no ensino fundamental e os indicadores média de hora-aula diária, percentual de alunos no turno noturno e idade mediana, constatou-se que a melhor transformação para reduzir a assimetria foi a aplicação do logaritmo.

Observou-se a ocorrência de assimetria à direita causada pela presença de valores extremos nas variáveis quantidade de ar-condicionado e mimeógrafo. Tomemos, por exemplo, a quantidade de aparelhos de ar-condicionado existente na escola: apenas 1% delas possui mais de 6 aparelhos, sendo que o máximo observado foi 94 aparelhos em uma escola. Nestes casos optou-se por utilizar uma variável indicadora, que vale 0 quando a escola não tem o aparelho e 1 caso contrário.

Durante a análise, verificou-se que as variáveis coletadas no Censo Escolar apresentam escalas extremamente diversificadas. Por exemplo, enquanto a variável “quantidade de alunos matriculados” assume valores entre 1 e mais de 10.000, o valor máximo do indicador “computador por aluno” é 4,3. Esta característica dos dados pode prejudicar ACP, como por exemplo, causar distorções no cálculo da fidedignidade dos componentes e também no cálculo dos escores finais. Variáveis com escalas de maior ordem de grandeza têm maior influência no valor final do escore calculado, sem que necessariamente exista maior carga associada. Desta forma, as variáveis quantitativas foram padronizadas para evitar tais efeitos.

### 3.2. Fatorabilidade da matriz de correlações

Uma vez definido o escopo das variáveis do questionário do Censo Escolar que farão parte da ACP, eliminados os *outliers* e aplicadas as transformações que permitiram adequar a distribuição das variáveis, avaliou-se a fatorabilidade da matriz de correlações. Este passo consiste em averiguar se a estrutura interna do conjunto de dados se adequa aos propósitos da ACP, ou seja, se na amostra ou censo em análise, as variáveis observadas mantêm algum nível de correlação entre si que justifique o intento de representá-las por meio de componentes.

Para avaliar a fatorabilidade da matriz de correlação observou-se o valor das correlações bivariadas e o critério de Kaiser-Meyer-Olkin – KMO. O critério KMO calculado para os dados vale 0,95, maior que o valor mínimo de 0,60 sugerido por Tabachnick e Fidell (2001). Tal resultado sinaliza a existência de uma estrutura de componentes principais entre as variáveis consideradas. A análise da matriz de correlações das 54 variáveis mostrou que 38,6% das correlações entre variáveis são maiores que 0,30, indício de fatorabilidade desta matriz.

Uma vez feitos os testes preliminares, passaremos às análises principais, que nos levará ao encontro da estrutura de componentes principais mais adequada para as variáveis do Censo Escolar voltados ao Ensino Fundamental.

### 3.3. Análises principais

Após a primeira etapa de refinamento dos dados, contamos com 54 variáveis e 172.460 escolas. Segue-se agora a determinação do número de componentes a extrair. Na ACP, esta decisão é de vital importância. Uma pesquisa com dados simulados (Zwick y Velicer, 1986) mostrou que o critério do autovalor maior do que 1 de Guttman-Kaiser (Guttman, 1954; Kaiser, 1960) é adequado apenas em 22% dos casos; o critério baseado no teste qui-quadrado de Bartlett (1950) é adequado em 30% dos dados; o critério baseado no teste scree, de Cattell (1966) é adequado em 57% dos casos; o critério de Velicer (1976), baseado nas correlações parciais, é adequado em 84% dos casos e o critério

baseado na Análise Paralela – AP, de Horn (1965), é adequado em 92% dos casos. Considerando estes resultados, iremos utilizar o critério do scree-plot aliado à análise paralela.

Na AP os autovalores de uma matriz de correlação de dados aleatórios são comparados com os da matriz empírica calculada pela análise dos componentes principais. Este procedimento de comparação dos autovalores, focaliza a magnitude de variância explicada pelos componentes. Segundo Reise, Walter e Comrey (2000), é pertinente reter um componente apenas na medida em que este explica maior variância do que o componente correspondente nos dados aleatórios. Assim, no momento em que o autovalor dos dados aleatórios é superior ao dos dados empíricos, não é mais adequado reter esse componente.

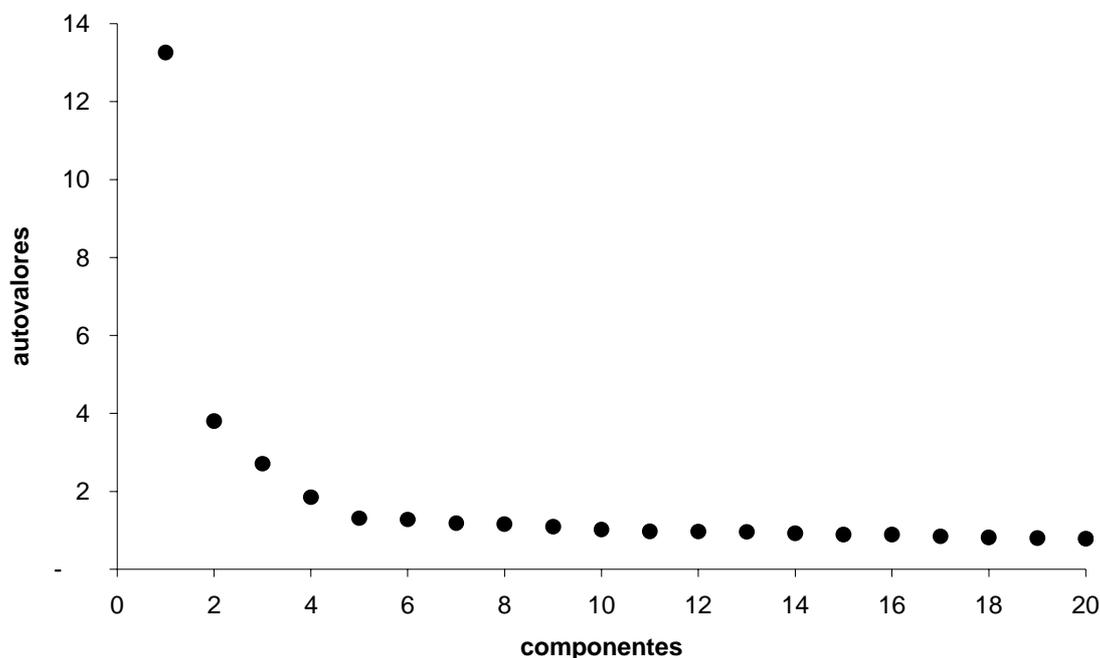
O procedimento desenvolvido em SPSS por Enzmann (2003) foi utilizado para identificar o valor dos autovalores aleatórios. A Tabela 2 mostra os autovalores empíricos – encontrados a partir da base de dados do Censo Escolar - e os autovalores aleatórios – encontrados à partir da macro. Como pode ser verificado, o componente 8 é o último em que os autovalores empíricos são superiores aos aleatórios. Já no componente 9, os valores empíricos são menores do que os valores aleatórios. Estes dados sugerem a existência de oito componentes.

O teste scree de Cattell (1966) é baseado no gráfico dos autovalores da matriz de correlação das variáveis em análise. Segundo este critério, a quantidade de componentes a reter é dada pelo ponto onde a curva apresenta a forma de 'cotovelo', significando que o ponto crítico se encontra no local do gráfico onde há uma forte mudança de inclinação da curva. A análise do scree-plot sugere o uso de 4 componentes.

TABELA 2. AUTOVALORES ALEATÓRIOS E EMPÍRICOS DOS 13 PRIMEIROS COMPONENTES

Componentes	Autovalores aleatórios	Autovalores empíricos	% variância explicada	% acum. de variância explicada
1	1,10	17,66	32,71	32,71
2	1,09	3,76	6,96	39,67
3	1,08	2,55	4,72	44,39
4	1,08	1,74	3,22	47,61
5	1,07	1,27	2,36	49,96
6	1,07	1,20	2,21	52,18
7	1,07	1,14	2,12	54,29
8	1,06	1,11	2,06	56,35
9	1,06	1,04	1,93	58,28
10	1,05	0,99	1,83	60,11
11	1,05	0,95	1,76	61,87
12	1,05	0,93	1,73	63,60
13	1,04	0,89	1,65	65,25

FIGURA 1. AUTOVALORES DA MATRIZ DE CORRELAÇÃO DAS 54 VARIÁVEIS EM ANÁLISE



Iniciamos a análise utilizando 8 componentes, com o objetivo de evitar a subestimação da quantidade de componentes a reter e priorizando a solução da Análise Paralela.

Outra decisão importante está relacionada ao uso da rotação na busca pela solução final de uma ACP. Na rotação ortogonal os fatores são mantidos não-correlacionados e na rotação oblíqua, em contraste, os fatores podem correlacionar (Laros, 2005). Foi utilizada a rotação Promax, pois trata-se da rotação oblíqua que apresentou melhor resultado final. Optou-se pelo uso de rotação oblíqua em função da existência de correlação entre os componentes finais.

Procurou-se obter uma solução simples, com os itens apresentando somente uma carga principal, para facilitar a interpretação. Isto pode ser conseguido retirando da estrutura de componentes principais as variáveis que apresentem (Laros e Puentes-Palácios, 2004):

1. Valor absoluto da carga principal na variável menor do que 0,32.
2. Diferença entre os valores absolutos das cargas principais de uma variável menor do que 0,10.
3. Ausência de similaridade entre o conteúdo do item e o domínio teórico do componente.

A aplicação iterativa desta regra até que nenhum item apresentasse qualquer uma das condições acima resultou em uma estrutura com três componentes (ver Tabela 3):

- Componente 1 – Infra-estrutura escolar e tecnológica.
- Componente 2 – Infra-estrutura básica e alimentar.
- Componente 3 – Condições favoráveis de ensino.

TABELA 3. MATRIZ DE CORRELAÇÃO ENTRE OS COMPONENTES

Componente	1	2	3
1	1,00	0,58	0,16
2	0,58	1,00	0,27
3	0,16	0,27	1,00

Como se pode observar na Tabela 3, que representa a matriz de correlação dos componentes extraídos, a correlação entre o Componente 1 e 2 é razoavelmente alta (0,58) a ponto de justificar o uso da rotação oblíqua, dentre as quais optamos pela Promax, por apresentar resultados mais claros do ponto de vista interpretativo.

A análise dos componentes principais deve ser complementada com a avaliação do quão confiável cada componente é em relação ao conceito que representa, o que pode ser feito por meio da verificação do coeficiente de fidedignidade. Essa medida permite mensurar a “consistência dos escores obtidos pelas mesmas pessoas quando elas são reexaminadas com o mesmo teste em diferentes ocasiões, ou com diferentes conjuntos de itens equivalentes, ou sob outras condições variáveis de exame” (Anastasi, 2000). O coeficiente de fidedignidade responde à pergunta “que porcentagem da variância do escore é atribuível a diferenças verdadeiras e não a erro?” (Cronbach, 1996). Utilizou-se o coeficiente de fidedignidade lambda 2 de Guttman ( $\lambda_2$ ).

Com o objetivo de facilitar as análises e padronizar os resultados obtidos nos três componentes encontrados, aplicou-se uma transformação de escala de tal forma que a distribuição geral dos escores passou a ter média 500 e desvio-padrão 100, escala também utilizada no GRE - Graduate Record Examination, e no SAT - College Board's Scholastic Aptitude Test, na análise dos escores dos três componentes encontrados,

O Componente 1: Infra-estrutura escolar e tecnológica ( $\lambda_2 = 0,91$ ) destaca-se por identificar, principalmente, os recursos de infra-estrutura tecnológicos disponíveis na escola, tais como quantidade média de computadores por aluno, existência de laboratório de informática, impressora. Também participam deste componente, embora com cargas menores, itens associados à infra-estrutura escolar geral da escola, tais como quadra de esporte, existência de merenda escolar, existência de laboratórios. As 4 escolas que possuem os itens descritos no componente são tipicamente escolas localizadas na zona urbana, que dispõem de recursos mais sofisticados para oferecer aos seus alunos. Reflete, ainda que de forma indireta, o perfil sócio-econômico da escola, uma vez que se tratam de artigos com curso relativamente elevado. Tem-se nesse componente, portanto, um indicador indireto da riqueza da escola.

TABELA 4. COMPONENTES PRINCIPAIS, COEFICIENTES (LOADINGS) DOS COMPONENTES APÓS ROTAÇÃO PROMAX, FIDEDIGNIDADE, COMUNALIDADE E CORRELAÇÃO ITEM-TOTAL

Item	Componentes			Comuna- lidade h <sup>2</sup>	Correlação Item- total
	1	2	3 <sup>(1)</sup>		
Coef. de fidedignidade ( $\lambda_2$ )	0,91	0,95	0,70		
Nº itens no componente	18	19	4		
Computador por aluno	<b>0,84</b>	0,03	-0,19	0,71	0,61
Laboratório de informática	<b>0,83</b>	-0,10	-0,03	0,60	0,58
Impressoras	<b>0,74</b>	0,18	0,07	0,77	0,80
Laboratório de ciências	<b>0,73</b>	-0,13	0,07	0,46	0,51
Internet	<b>0,73</b>	0,00	-0,02	0,53	0,59
Aparelho de fax	<b>0,72</b>	0,03	0,07	0,57	0,64
Máquina copiadora	<b>0,70</b>	0,05	-0,02	0,53	0,60
Merenda escolar	<b>-0,68</b>	0,10	0,46	0,53	0,27
Quadra de esporte coberta	<b>0,60</b>	-0,05	0,04	0,34	0,45
Retroprojektor	<b>0,57</b>	0,19	0,21	0,59	0,71
Ar-condicionado	<b>0,57</b>	-0,14	-0,06	0,25	0,32
Videoteca	<b>0,54</b>	-0,08	-0,05	0,25	0,35
Outros laboratórios	<b>0,54</b>	-0,16	0,01	0,22	0,30
Sala de TV e vídeo	<b>0,53</b>	0,18	0,00	0,42	0,57
Piscina	<b>0,49</b>	-0,12	-0,22	0,21	0,20
Cantina	<b>0,48</b>	-0,05	0,06	0,22	0,37
Almoxarifado	<b>0,43</b>	0,28	0,03	0,40	0,58
Sala de leitura	<b>0,42</b>	0,07	-0,06	0,21	0,36
Liquidificador	-0,06	<b>0,82</b>	0,03	0,63	0,64
Energia pública	-0,12	<b>0,81</b>	-0,17	0,52	0,50
Sanitário dentro do prédio	-0,14	<b>0,81</b>	-0,19	0,50	0,48
Geladeira	0,06	<b>0,78</b>	-0,09	0,63	0,65
Coleta pública de lixo	0,18	<b>0,74</b>	-0,08	0,70	0,73
Freezer	0,03	<b>0,71</b>	0,16	0,63	0,69
Água pública	0,15	<b>0,68</b>	-0,03	0,59	0,68
Secretaria	0,16	<b>0,66</b>	0,02	0,59	0,61
Antena parabólica	-0,08	<b>0,66</b>	0,32	0,59	0,69
Alunos matriculados no EF	0,11	<b>0,59</b>	0,48	0,82	0,80
Aparelho de vídeo	0,38	<b>0,56</b>	0,13	0,76	0,84
TV	0,39	<b>0,55</b>	0,16	0,79	0,86
Sanitário fora do prédio	-0,31	<b>0,54</b>	-0,08	0,18	0,16
Cozinha	-0,23	<b>0,53</b>	-0,05	0,18	0,22
Diretoria	0,31	<b>0,53</b>	0,08	0,60	0,73
Depósito de alimentos	-0,26	<b>0,50</b>	0,30	0,31	0,32
Sala de professores	0,35	<b>0,50</b>	0,11	0,63	0,75
Ventiladores	0,35	<b>0,46</b>	0,06	0,54	0,69
Refeitório	0,16	<b>0,39</b>	-0,03	0,24	0,42
Idade mediana	0,20	-0,29	<b>0,84</b>	0,68	0,29
Distorção idade-série	-0,21	-0,57	<b>0,71</b>	0,75	-0,36
Alunos por turma	0,00	0,19	<b>0,64</b>	0,52	0,41
% alunos noturno	-0,13	0,12	<b>0,60</b>	0,39	0,24

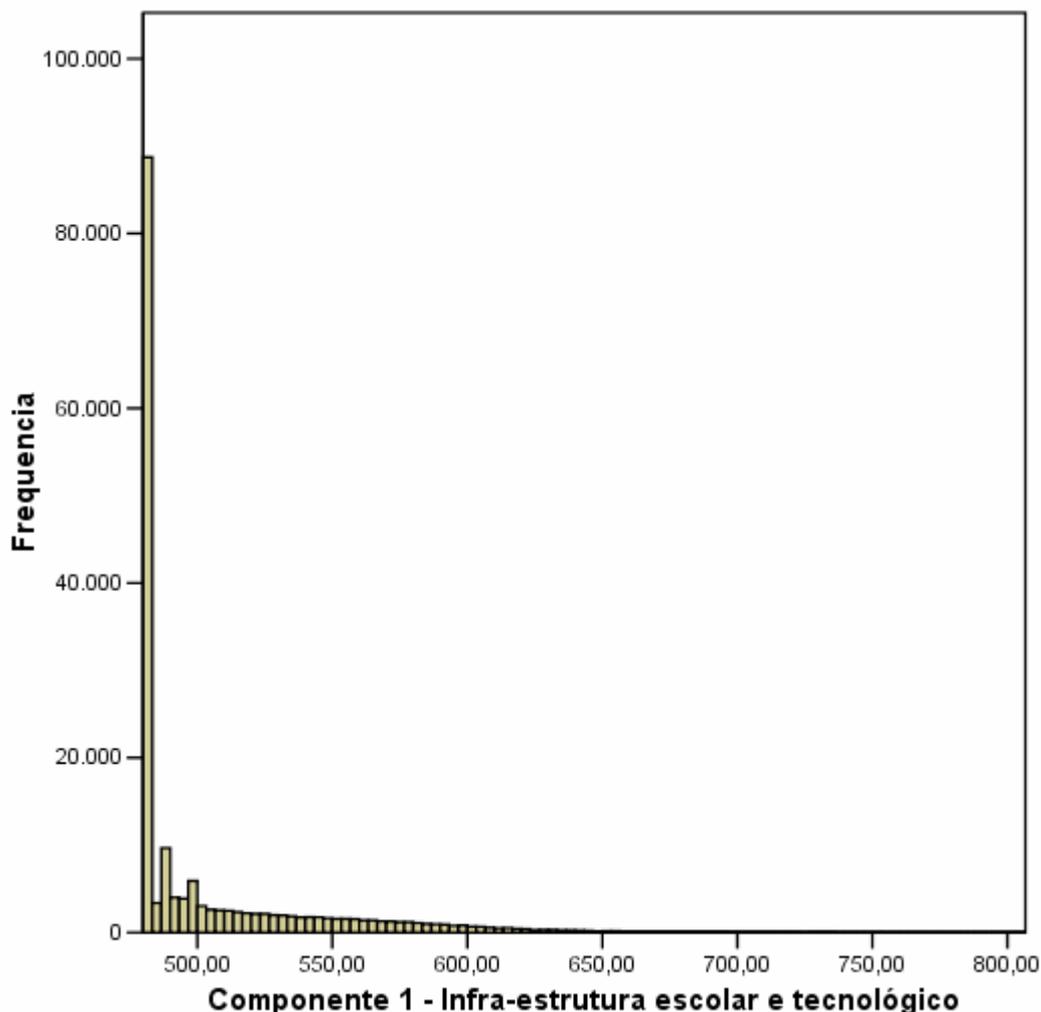
Nota: Componente 1: Infra-estrutura escolar e tecnológica; Componente 2: Infra-estrutura básica e alimentar; Componente 3: Condições favoráveis de ensino.

(1) A interpretação dos escores das escolas no Componente 3: "Condições favoráveis de ensino" deve ser feita da seguinte forma: quanto menor o escore, mais favoráveis as condições de ensino na escola.

A distribuição do escore do Componente 1 tem forte assimetria à direita, demonstrando que as escolas brasileiras são, via de regra, carentes em relação aos recursos aqui mencionados. São poucas as

escolas que possuem todos itens e a observação detalhada dos dados indica que, dentre as escolas que possuem alguns itens, os mais comuns são o retroprojektor e o aparelho de fax.

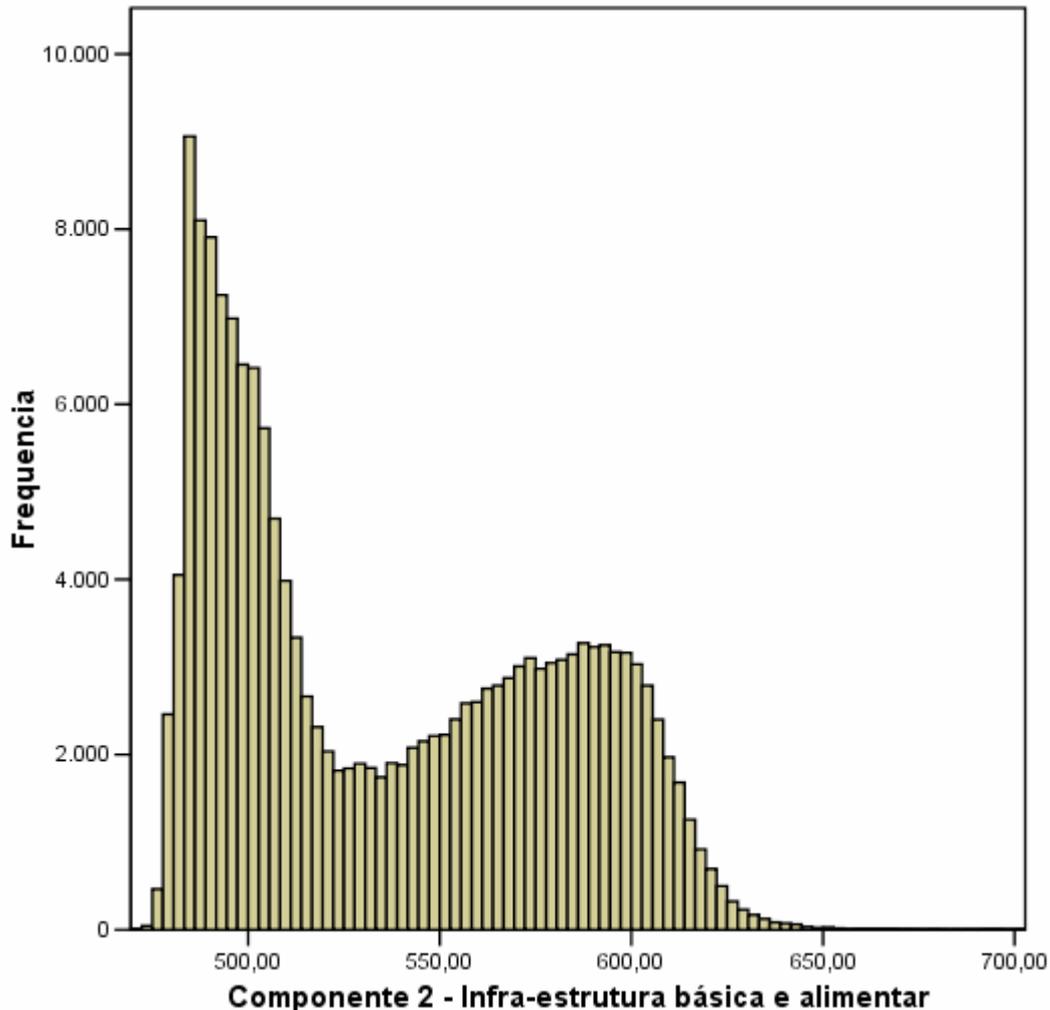
FIGURA 2. HISTOGRAMA DOS ESCORES DO COMPONENTE 1



O Componente 2: Infra-estrutura básica e alimentar ( $\lambda_2 = 0,95$ ) recebe esse nome por ser formado por itens que indicam a existência de infra-estrutura básica (energia pública, sanitários, coleta pública de lixo, água pública) bem como por itens que indicam quais recursos estão disponíveis na escola para fornecer ou facilitar o fornecimento de alimentos aos alunos: liquidificador, geladeira, freezer, cozinha, depósito de alimentos e refeitório. Quanto maior o escore da escola nesse componente, melhor a infra-estrutura disponível na escola, tanto do ponto de vista básico quanto para facilitar a disponibilização de refeições ao seu corpo discente.

A distribuição dos escores do Componente 2 mostra que uma grande parte das escolas brasileiras não dispõe de todos recursos de infra-estrutura básica e alimentar considerados, uma vez que há assimetria à direita. Verifica-se dois grandes focos de concentração, caracterizando uma divisão entre as escolas de ensino fundamental: as que contam apenas com os recursos básicos e as que dispõem em suas dependências de recursos destinados à melhor oferta alimentar.

FIGURA 3. HISTOGRAMA DOS ESCORES DO COMPONENTE 2



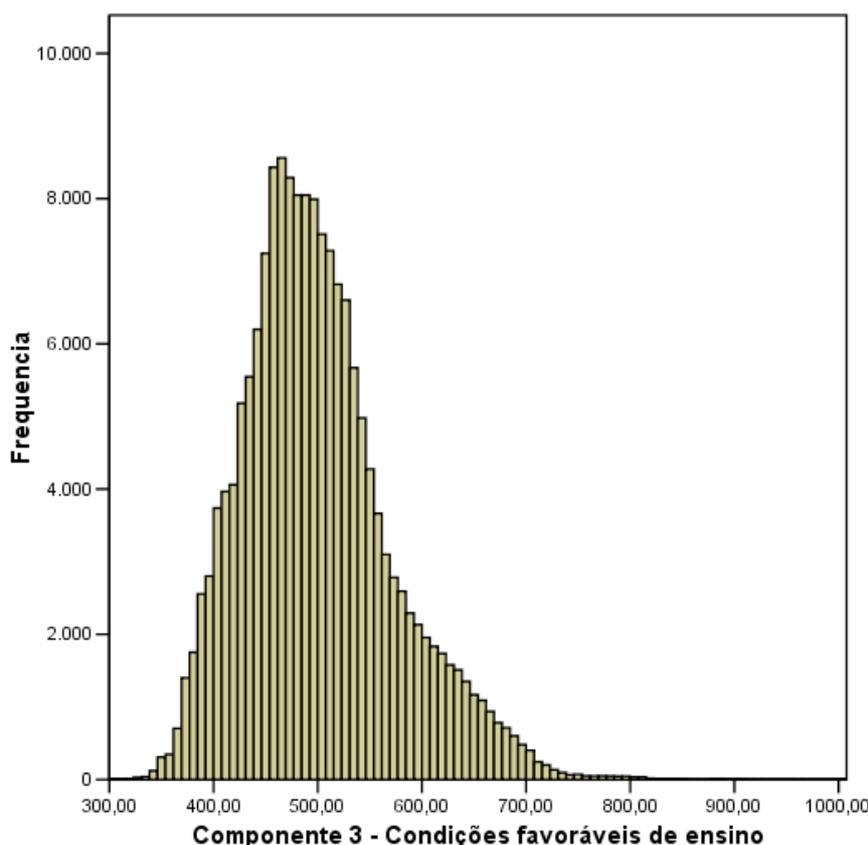
O componente “Condições favoráveis de ensino” ( $\lambda_2 = 0,70$ ), é formado por quatro indicadores derivados do Censo Escolar e reflete três aspectos da composição da escola: a possível inadequação da idade dos alunos que cursam o ensino fundamental, o tamanho médio da turma, e o horário de frequência às aulas. Estes aspectos podem estar relacionados ao bom desempenho escolar, principalmente em estabelecimentos que adotam a promoção seriada, uma vez que o escore é fortemente influenciado pela adequação idade-série. No sistema seriado convencional, um aluno só é considerado apto a frequentar a série seguinte se tiver sido aprovado na série que frequenta, de maneira que pode-se associar a idade do aluno ao seu rendimento escolar: um estudante com sete anos de idade deve, idealmente, estar cursando a 1ª série do ensino fundamental, e assim por diante. O fato de o aluno ter idade superior à adequada para a série pode ser causado pela repetência ou desistência em anos anteriores, ou ainda pelo ingresso com atraso do aluno na escola. Ocorre que, o art. 23 da Lei de Diretrizes e Bases da Educação de 1996 introduziu a autonomia aos estabelecimentos de ensino no que diz respeito à escolha da organização dos tempos escolares. Hoje vivemos uma realidade em que as escolas adotam, principalmente, três critérios de promoção: seriada, em ciclos e mista. Estados como Rio de Janeiro, São Paulo, Porto Alegre, Belém e Belo Horizonte, implantaram reformas

substanciais na reorganização dos tempos escolares em suas redes de ensino (Negreiros, 2005; Franco, 2004). Assim, é necessário analisar esse componente com cautela nos estados e municípios onde a seriação deixou de ser o modelo usual. Para algumas unidades da federação, a distorção idade-série pode ser reduzida devido à aprovação automática dentro de um ciclo, e não por que o ensino está mais eficiente. Isso pode, por exemplo, ter influenciado na boa posição de São Paulo no escore, se comparado com as demais UF.

Outros aspectos chamam a atenção no componente “Condições favoráveis de ensino”. Alunos que estudam a noite em geral têm menos tempo para estudar em casa. Conclui-se que um maior escore nesse componente corresponderá a piores condições de ensino, pois indica escolas onde os alunos têm idade superior à esperada para que se frequente o ensino fundamental, as turmas são excessivamente grandes e há muitos alunos frequentando o turno noturno. É importante destacar que a interpretação do escore deste componente se dá de forma inversa aos demais: quanto maior o escore, mais distante está a escola da situação ideal, onde os alunos frequentam a série adequada à sua idade, as turmas são pequenas e as aulas têm duração suficiente.

Em relação ao Componente 3 – Condições favoráveis de ensino, este tem a particularidade de obedecer a uma escala inversa aos demais fatores, no sentido de que quanto menor o escore neste componente, melhor as condições de ensino. Em geral as escolas brasileiras do Ensino Fundamental têm condições próximas, com poucas escolas se destacando como excelentes ou como ruins. A distribuição do escore é levemente assimétrica à direita, fato que reforça a afirmação anterior.

FIGURA 4. HISTOGRAMA DOS ESCORES DO COMPONENTE 3



#### 4. DISCUSSÃO E CONCLUSÃO

O Censo Escolar é um instrumento de coleta de dados capaz de agregar muitas informações de valor à análise da educação brasileira. Neste estudo procuramos principalmente sintetizar de forma clara e precisa as dimensões abordadas no levantamento anual, que cumpre um papel importante nas determinações das políticas públicas, além de servir de subsídio para vários estudos. Citamos com exemplo Sampaio *et al.* (2002), que analisa a questão do sincronismo idade/série utilizando intensamente os dados do Censo Escolar de 1996 a 2001. Gatti (1004) faz uma revisão dos estudos quantitativos em educação e, dentre os estudos citados, percebe-se a presença indireta do Censo Escolar – é o caso, por exemplo, dos estudos envolvendo o Saeb e a educação infantil.

A ACP mostra que uma grande colaboração do Censo Escolar está na caracterização das escolas quanto à infra-estrutura oferecida pelo estabelecimento de ensino. Outra colaboração está no diagnóstico das condições de ensino com base nos indicadores educacionais (Inep, 2004), aqui representados por um único fator, cujo conteúdo aborda principalmente aspectos etários do alunado e relacionados ao horário de estudo.

Além de resumir o conjunto de variáveis coletados no Censo Escolar, a ACP nos possibilita realizar o cálculo dos escores de cada escola em relação aos componentes. A análise da distribuição dos escores dos componentes principais permite descrever as características das escolas brasileiras de ensino fundamental em relação aos construtos mensurados. Seu uso vai desde o enfoque localizado, em que o interesse estaria em municípios ou mesmo escolas, até a comparação entre as unidades da federação e regiões administrativas do Brasil. Neste estudo optou-se por uma análise geral de cada componente, seguida da comparação entre as unidades da federação brasileira.

A análise dos escores do componente principal pode ser enriquecida comparando as unidades da federação. Stoneberg (2005) sugere um procedimento para este tipo de comparação: calcular os intervalos de confiança dos escores de cada UF e comparar os limites superiores e inferiores, evitando o uso das estimativas pontuais. A desvantagem em utilizar estimativas pontuais do escore para elaborar listas de classificação está na possibilidade de classificar um estado como melhor que outro quando seus escores possuem valores muito próximos, de maneira que a diferença não seja significativa a um dado nível de confiança.

Para evitar dificuldades deste tipo, Stoneberg (2005) sugere algo simples no seu estudo dos Estados Unidos, que pode ser estendido ao caso brasileiro: escolher um estado e comparar os outros em relação ao estado escolhido, classificando-os em significativamente maior que o estado escolhido, sem diferença significativa ou significativamente menor que o estado escolhido. Neste estudo vamos fixar Pernambuco como o estado base para comparação com as demais unidades da federação brasileira, pois em todos componentes esse estado se posiciona aproximadamente equidistante entre os maiores e menores escores por unidade da federação.

Tratando-se da Infra-estrutura escolar e tecnológica, observa-se que Pernambuco não é uma das UF que dispõe de maior abundância. Seu posicionamento em relação às demais UF do Brasil indica ser este um estado com condições superiores a 13 estados, porém com possibilidades de grande ampliação de recursos voltados para o atendimento das necessidades escolares e de tecnologia atuais. Levando em conta que o componente, como um todo, apresenta assimetria à esquerda, deduz-se que o estado em análise conta com boa infra-estrutura escolar, se comparado com aqueles onde a situação é mais precária.

TABELA 5. ORDENAMENTO DOS ESCORES MÉDIOS POR COMPONENTE EM RELAÇÃO AO ESTADO DE PERNAMBUCO

Unidade da federação	N	Componente 1			Componente 2			Componente 3		
		LI	LS	PE	LI	LS	PE	LI	LS	PE
Acre	1.553	487,77	487,83	(+)	506,45	506,57	(+)	510,24	510,40	(-)
Alagoas	3.500	491,05	491,09	(+)	529,68	529,73	(-)	529,98	530,08	(-)
Amazonas	4.679	492,71	492,74	(+)	509,53	509,57	(+)	527,62	527,69	(-)
Amapá	643	496,46	496,69	(=)	538,02	538,36	(-)	490,53	491,00	(+)
Bahia	24.418	490,30	490,31	(+)	517,25	517,26	(+)	529,18	529,19	(-)
Ceará	10.307	500,36	500,38	(-)	534,01	534,02	(-)	494,67	494,70	(+)
Distrito Federal	733	564,50	564,81	(-)	582,11	582,26	(-)	481,77	482,21	(+)
Espírito Santo	3.233	505,73	505,79	(-)	540,86	540,93	(-)	460,42	460,51	(+)
Goiás	4.112	512,21	512,25	(-)	556,19	556,23	(-)	496,83	496,94	(+)
Maranhão	12.763	486,23	486,24	(+)	510,89	510,90	(+)	529,81	529,84	(-)
Minas Gerais	13.378	513,02	513,04	(-)	550,93	550,95	(-)	485,08	485,11	(+)
Mato Grosso do Sul	1.117	545,58	545,77	(-)	585,86	585,97	(-)	509,77	510,15	(-)
Mato Grosso	2.557	508,90	508,97	(-)	542,67	542,77	(-)	491,68	491,82	(+)
Pará	12.312	487,43	487,44	(+)	509,24	509,25	(+)	524,40	524,42	(-)
Paraíba	6.630	491,92	491,93	(+)	525,50	525,52	(+)	506,57	506,63	(-)
<b>Pernambuco</b>	<b>10.643</b>	<b>496,62</b>	<b>496,63</b>	<b>(=)</b>	<b>528,39</b>	<b>528,41</b>	<b>(=)</b>	<b>501,64</b>	<b>501,68</b>	<b>(=)</b>
Piauí	7.428	487,41	487,43	(+)	512,70	512,72	(+)	511,87	511,91	(-)
Paraná	6.730	531,69	531,72	(-)	565,14	565,17	(-)	465,78	465,82	(+)
Rio de Janeiro	8.016	540,46	540,49	(-)	575,47	575,49	(-)	480,85	480,90	(+)
Rio Grande do Norte	3.924	496,46	496,49	(+)	537,41	537,46	(-)	486,58	486,67	(+)
Rondônia	2.453	493,34	493,39	(+)	515,81	515,90	(+)	471,91	472,02	(+)
Roraima	647	490,69	490,86	(+)	519,50	519,86	(+)	461,33	461,70	(+)
Rio Grande do Sul	8.591	520,93	520,96	(-)	556,84	556,87	(-)	453,47	453,51	(+)
Santa Catarina	4.498	522,37	522,42	(-)	557,03	557,07	(-)	449,34	449,39	(+)
Sergipe	2.255	493,60	493,66	(+)	529,45	529,53	(-)	517,32	517,48	(-)
São Paulo	13.155	559,48	559,50	(-)	582,79	582,80	(-)	479,85	479,88	(+)
Tocantins	2.185	496,70	496,77	(-)	524,84	524,95	(+)	505,57	505,72	(-)

Legendas: N: número de escolas; LI: Limite inferior do intervalo a 99% de confiança; LS: Limite superior do intervalo a 95% de confiança; PE: comparação com o estado de Pernambuco; (+): diferença significativamente superior; (=) não há diferença significativa; (-) diferença significativamente inferior.  
 Nota: Componente 1 – Infra-estrutura escolar e tecnológica; Componente 2: Infra-estrutura básica e alimentar; Componente 3: Condições favoráveis de ensino

Em relação à “Infra-estrutura básica e alimentar”, observa-se que o estado de Pernambuco conta relativamente com menos recursos que 10 UF. Reflete-se aí o posicionamento de Pernambuco à esquerda da distribuição, ou seja, com poucos recurso desse gênero, se comparado com outros estados do Brasil.

No componente “Condições favoráveis de ensino” observa-se que Pernambuco é superado por 15 outras UF, tipicamente as com melhores indicadores escolares, recordando que nesse caso, quanto menor o escore, melhores condições de ensino.

Como resultado geral desse estudo concluímos que as informações disponíveis no Censo Escolar podem ser utilizadas como meio de conhecimento das escolas brasileiras no que diz respeito a seus aspectos físicos e às características de seus alunos. Tais informações podem ser úteis em várias análises, servindo como covariáveis nas análises de proficiência que utilizam as bases de dados do Saeb, por exemplo. Permitem também identificar o perfil da escola brasileira, destacar carências e

localizar regiões geográficas onde exista maior necessidade de atendimento por parte da administração em aspectos com infra-estrutura básica e tecnológica. Fazer uso inteligente dos dados coletados anualmente por meio do Censo Escolar significa conhecer melhor a realidade de nossas escolas e possibilitar a atuação de forma mais precisa.

## REFERENCIAS BIBLIOGRÁFICAS

- Anastasi, A. y Urbina, S. (2000). *Testagem psicológica*. Porto Alegre: ArtMed Editora.
- Bartlett, M.S. (1996). Test of significance in factor analysis. *British Journal of Psychology*, 3, pp. 77-85.
- Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, pp. 245-276.
- Cronbach, L.J. (1996). *Fundamentos da testagem psicológica*. Porto Alegre: Artes Médicas.
- Diniz, E.M.V. (1999). O Censo Escolar. *Revista Brasileira de Estudos Pedagógicos*, 80(194), pp. 156-163.
- Enzmann, D. (1997). RanEigen: A program to determine the parallel analysis criterion for the number of principal components. *Applied Psychological Measurement*, 21, pp. 232.
- Enzmann, D. (2003). *RanEigen 2.0*. Retirado do <http://www.kfn.de/softwareenzmann.html>.
- Franco, C. (2004). Ciclos e letramento na fase inicial do ensino fundamental. *Revista Brasileira de Educação, Rio de Janeiro*, 25, pp. 30-38.
- Gatti, B.A. (2004). Estudos quantitativos em educação. *Educação e Pesquisa*, 30(1), pp. 11-30.
- Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, 19, pp. 149-162.
- Hopkins, K.D. (1998). *Educational and psychological measurement and evaluation*. Needham Heights, MA: Allyn and Bacon.
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, pp. 179-185.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2000). *Geografia da Educação Brasileira*. Brasília, DF: Autor.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2001). *Sinopse Estatística da Educação Básica 2001*. Brasília, DF: Autor.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2003). *Mapa do analfabetismo no Brasil*. Brasília, DF: Autor.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2004). *Dicionário de indicadores educacionais – fórmulas de cálculo*. Brasília, DF: Autor.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2006). *O que é o Censo Escolar e qual o seu objetivo?*. Retirado em 7-3-2006 do <http://www.inep.gov.br/basica/censo/escolar>.
- Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, pp. 141-151.
- Kvanli, A.H., Guynes, C. S. y Pavur, R. J. (1996). *Introduction to Business Statistics. A computer Integrated, Data Analysis Approach*. St. Paul Minneapolis: West Publishing Company.
- Laros, J.A. y Puente-Palacios, K.E. (2004). Validação cruzada de uma escala de clima organizacional. *Estudos de Psicologia*, 91, pp. 113-119.
- Laros, J.A. (2005). O uso da análise fatorial: algumas diretrizes para pesquisadores. En L. Pasquali, *Análise Fatorial Para Pesquisadores*. Brasília: LabPAM.
- Negreiros, P.R.V. (2005). Séries no ensino privado, ciclos no público: um estudo em Belo Horizonte. *Cadernos de Pesquisa*, 35(125). pp. 181-203.

- Osborne, J.W. (2002). Notes on the use of data transformations. *Practical Assessment, Research & Evaluation*, 86. Retirado em 11/11/2005 de <http://PAREonline.net>.
- Reise, S.P., Waller, N.G. y Comrey, A.L. (2000). Factor analysis and scale revision. *Psychological Assessment*, 12, pp. 287-297.
- Sampaio, C.E.M., Sousa, C.P., Andrade, F.C., Santos, J.R.S., Pereira, J.V., Costa, J.R., Brant, L.L.N.A.O., Oliveira, M.R., Campos, R.M.R. y Oliveira, V.N. (2002). Sincronismo idade/série: um indicador de produtividade do sistema educacional brasileiro. *Inep - Textos para discussão*, 11.
- Stoneberg, B.D. (2005). Please don't use NAEP scores to rank order the 50 states. *Practical Assessment, Research & Evaluation*, 910. Retirado em 15/11/2005 de <http://PAREonline.net>.
- Tabachnick, B.G. y Fidell, L. S. (2001). *Using multivariate statistics*. MA: Allyn and Bacon.
- Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, pp. 321-327.
- Zwick, R. y Velicer, W.F. (1986). Comparison of five rules for determining the number of componentes to retain. *Psychological Bulletin*, n. 99, pp. 432-442.